TECHNIQUES OF MODEL BUILDING FOR DEVELOPING ECONOMIES

Lawrence R. KLEIN
and
Stefan SCHLEICHER

Forschungsbericht/Research Memorandum
No. 91

April 1975

INDEX

1.0. DATA BASE

The quality of final products is obviously closely related to their basic material contents. Occasionally a highly skilled craftsman can overcome raw material deficiencies and turn out an acceptable final product in spite of material input deficiencies, but this is unusual and certainly not dependable; therefore, we believe that we can say that no econometric model can be good if it does not rest on a good data base. Perhaps it would be better to say that the quality of the data base will restrain the quality performance of any model built on this base and it is a challenge to the econometrician to build models that serve end use needs as well as possible, up to certain minimum standards, given the data base.

Even in the best of cases, in dealing with models for the most advanced industrial nations, there are serious deficiencies in basic quality of economic data. Such data are "noisy", i.e. they are subject to direct observational or conceptual error. But in the cases of developing nations, countries whose history of industrial development on a modern scale is relatively brief, we find much more serious problems of data deficiencies even to the point of completely discouraging many investigators from attempting to model the implied development process through econometric methods.

Let us start out this discussion on a relatively optimistic note, namely, that although there are serious data problems facing the econometric analysis of the developing economy, we can successfully overcome the callenge and build econometric models of the

LDC that are useful and capable of providing economic insights that we would not otherwise have. Part of the problem in specifying and estimating econometric models of the LDC is in recognizing the full extent of data deficiencies at the outset and designing model and method so as best to cope with the situation. This makes the model building problem different for the LDC. It should also govern our aspirations; nevertheless, the rewards should turn out to be significant.

There are two dimensions of data deficiency that we must recognize at the outset:

1. Data quality-error of observation and lack of availability of key data.
2. Sparseness of statistical samples.

The quality aspect means that we shall have trouble measuring the things that we think we are measuring or intending to measure. These will take the form both of <u>bias</u> and <u>variance</u>. Estimation methods that account for "errors-in-variables" will be in order. Quality deficiencies also mean that data required to measure some sophisticated concepts in economic analysis will not be available at all if we impose minimum standards. The concepts of capital consumption, capital stock, disguised unemployment, effective interest rate, inventory investment, and others should be deemphasized or deleted from explicit display in model construction for economies where it is plainly not going to be possible to secure data on measurement of these attributes of a system. The concepts still exist and play important roles in the working of the economies concerned, but they defy explicit numerical measurement; so we must build our analysis in such a way as not to depend on them explicitly.

The sparseness of data refers to the idea that there may be many missing observations. Statistical time series will be short, in the sense that they will start only from a fairly recent origin. Also there may be gaps. Measurement may be only biennial, quinquennial or decadal. Some time periods, on a purely unsystematic basis, may be entirely missing from records. The scope of measurement is likely to be limited - covering only some of the magnitudes of interest to the econometrician. Small sample methods will be the order of the day in coping with a shortness of "degrees of freedom" in the statistical sense.

To be very practical and realistic, it should be noted that statistical records of developing countries are generally available only for the period since the end of World War II, and usually only since a period lying somewhere between 1955 and 1960. They are most often available as annual time series. This clearly means that no more than twenty statistical observations or data points are going to be available in most cases. This is not hopeless, however, when we recognize that the first attempts at modeling the U.S. economy started out with approximately the same number of data points.[1)] The "interwar" sample (between World Wars I and II) formed the original base, in analogy to our present "postwar" sample (the period since World War II or in some cases, the Korean War).

---

1) J. Tinbergen, Statistical Testing of Business-Cycle Theories, Vol. I, II, (Geneva: League of Nations, 1939) L.R. Klein, Economic Fluctuations in the United States, 1921-1941, (N.Y.:Wiley,1950)

Samples of twenty data points are barely sufficient for estimation of stable statistical relationships, but they do appear to provide enough information to allow useful estimates to be made. A major problem in dealing with small samples, such as twenty annual observations for a macroeconomy, is that many explanatory variables is each separate relationship will move in common trend patterns. For sound statistical estimates we need variability, and <u>independent</u> variability among different explanatory variables; otherwise we encounter the familiar difficulties of multicollinearity. In a linear relationship - one of several in a complete system - we shall want to estimate the coefficients (αi)

$$Y_t = \sum_{i=1}^{n} \alpha_i X_{it} + e_t$$

where

Y_t = dependent variable

X_{it} = explanatory variables, i = 1,2, ...n

e_t = residual error

If the different X_{it} are interrelated, we have some degree of collinearity. Our problem is that we wish to estimate the separate contributions of each X_{it} in explaining the historical movement of Y_t, but to the extent that the X_{it} move together in the same pattern, their separate contributions are obscured, in a statistical sense.

The crucial aspects of the problem can be readily illustrated in the simple case, n = 2.

$$Y_t = \alpha_1 X_{1t} + \alpha_2 X_{2t} + e_t$$

The constant term of the linear equation is deleted, on the assumption that all variables are adjusted to have zero means. The least-squares estimation equations for α_1, α_2 are the wellknown "normal" equations

$$\sum_{t=1}^{T} Y_t X_{1t} = \alpha_1 \Sigma X_{1t}^2 + \alpha_2 \Sigma X_{1t} X_{2t}$$

$$\sum_{t=1}^{T} Y_t X_{2t} = \alpha_1 \sum_{t=1}^{T} X_{1t} X_{2t} + \alpha_2 \sum_{t=1}^{T} X_{2t}^2$$

These equations are more revealing, for the problem at hand, if we standardize the variables Y_1, X_{1t} and X_{2t} by dividing each by its own standard deviation; thus the standardized variable will have unit standard deviation. The standardized form of the equation to be estimated is

$$\frac{Y_t}{\sigma_y} = \alpha_1' \frac{X_{it}}{\sigma_1} + \alpha_2' \frac{X_{2t}}{\sigma_2} + e_t'$$

The corresponding "normal" equations are

$$r_{1y} = \alpha_1' r_{11} + \alpha_2' r_{12}$$

$$r_{2y} = \alpha_1' r_{12} + \alpha_2' r_{22}$$

Instead of using moments of the variables as coefficients and left-hand-side quantities in these equations, we use correlation coefficients. We know that

$$r_{11} = 1, \quad r_{22} = 1$$

The solution for estimates of α_1 and α_2, the transformed parameters, are given by

$$\text{est. } \alpha_1' = \frac{\begin{vmatrix} r_{1y} & r_{12} \\ r_{2y} & 1 \end{vmatrix}}{\begin{vmatrix} 1 & r_{12} \\ r_{12} & 1 \end{vmatrix}}$$

$$\text{est. } \alpha_2' = \frac{\begin{vmatrix} 1 & r_{1y} \\ r_{12} & r_{2y} \end{vmatrix}}{\begin{vmatrix} 1 & r_{12} \\ r_{12} & 1 \end{vmatrix}}$$

it is now evident that as $r_{12} \to 1.0$, indicating that the two explanatory variables have moved together, we find that the denominators for estimates of α_1', α_2' each approach zero. It is also the case that the numerators approach zero. This is because $r_{1y} \to r_{2y}$; i.e. as X_{1t} and X_{2t} approach perfect correlation with each other, they approach the <u>same</u> correlation with Y_t. It is only in the limit as $r_{12} \to 1$ that the full difficulties of collinearity arise. In that case, each of the two estimates becomes 0/0, an indeterminate form. Where collinearity is strong, but not at the limit, the estimates are very sensitive to computation error or observation error.

The estimates then exhibit statistical instability, not complete indeterminacy. The exposition is presented here for the simplest case of ordinary least squares regression with two explanatory variables. There are two significant directions in which all these ideas generalize. The same kinds of problems arise for multivariate equations with more than two explanatory variables and also with a situation of general linear dependence among the explanatory variables, or a subset of them. It is not purely a case of simple bivariate correlation between pairs of explanatory variables.

Secondly, equally grave problems arise for other estimation methods applied to individual equations that form part of complete systems. In fact, it is shown elsewhere that more sophisticated methods of estimation are more sensitive than OLS methods to the presence of multicolliniarity.[1)]

This is the problem of collinearity, what can be done about it? There is little that can be done, using only the information at hand. The only satisfactory approach is to collect added information. Two types are relevant. On the one hand, the investigation should seek some a priori (non sample) information about α_1 or α_2. If either can be fixed or partly so, in advance, then only one parameter needs to be estimated from sample data, and the collinearity problem should not arise. The prior information can be taken from abstract economic analysis (optimization theory), from engineering information, from knowledge of economic institutions, etc.

1) L.R. Klein and M. Nakamura, "Singularity in the Equation Systems of Econometrics: Some Aspects of the Problem of Multicollinearity." International Economic Review, 3 (Sept. 1962), 274-99.

The other way of providing new information is to draw a fresh sample - not necessarily the same kind of time-series sample, for that is usually not possible. A more promising line is to draw, if possible, a cross-section sample - one that varies across observation units at a given time point and not across time points. This is not an empty suggestion. We often have samples of individual family budgets or individual behavior of producing units, such as the firm.

In a cross-section sample, as in the case of a study of household family budgets, we have income and expenditure data for individual units during a given span of time, often as short as a few weeks or months. From these data we can estimate Engel curves with income effects on spending, for a given set of prices that prevail for all families during the period of the sampling for the budget investigation. The point to be emphasized about this sample is that it effectively breaks up the collinearity between prices and income variation. It also enlarges the scope of the sample because observations on several hundred or thousand household units make up the sample points. In addition to sample enlargement and the break up of collinearity, the Engel elasticity or coefficient should give a good indication of long run behavior.

It is by no means unusual to find usable family budget studies in developing countries. A substantial investigation of Colombian family budgets from four areas has just been thoroughly investigated in a doctoral dissertation.[1)] This particular study is remarkable

1) Howard Howe, Estimation of the Linear and Quadratic Expenditure Systems: A Cross-Section Case for Colombia, Ph.D. thesis, University of Pennsylvania

in showing how such data can be collected, processed, and used in statistical estimation of economic relationships for developing countries. Moreover, the estimated relationships were sophisticated sets of complete demand systems, the linear expenditure systems (LES) and the quadratic expenditure system (QES).

There is not, however, a straightforward gain in supplementing sample information with cross-section data. Complicating factors are the introduction of new sources of variation in the form of particular variables that influence individual family behavior. These variables tend to get averaged out and change imperceptibly in time series samples, but they vary widely and are of great importance for explaining individual behavior. In the case of family expenditure behavior such particular variables as family size, family composition, age of head, region of residence, occupation of head, and others are important. The complicating feature of using such data for estimation of income effects is that the investigation must adjust the sample in order to standardize families for these particular variables or must introduce them in larger multivariate relationships.

A technique for combining information from cross-section and time series data is to estimate the income coefficient from the former and the price coefficient from the latter. If an overall demand function for a major expenditure class (food, clothing, shelter, fuel, services, or durables) takes the log linear form

$$\log X_t = \alpha_0 + \alpha_1 \log \frac{P_{xt}}{P_t} + \alpha_2 \log \frac{Y_t}{P_t} + e_t$$

X_t = real market demand for a commodity

P_{xt} = price of X

P_t = general price level

Y_t = nominal aggregate income level

e_t = random error

We can work out direct procedures for using both data types. To estimate α_2 first from a family budget sample, we specify the individual equation

$$\log P_{xto} X_{ito} - \log P_{xto} = \alpha_o' + \alpha_1 \log \frac{P_{xto}}{P_{to}} + \alpha_2 \log \frac{Y_{ito}}{P_{to}} + e_{ito} + \sum_{j=1}^{n} \beta_j Z_{ijto}$$

$$i = 1,2 \ldots, N$$

Data for individual families will be available for expenditures ($P_{xto} X_{ito}$) and not generally for real demand. The expenditure and income variables have i subscripts because they are associated with the individual. The same is true of the error term and the particular "extra" variables Z_{ijto}. In contrast, the price variables P_{xto} and P_{to} do not have i subscripts. They are assumed to be the same for all individuals. Since the sample is taken at a given time point, the subscript variable t is fixed at t_o.

Using least squares regression, to be explained in a later section, or other methods, estimates of α_2 could be obtained from a sample of N observations. The estimate will be

$$\log P_{xto} X_{ito} = a + a_2 \log Y_{ito} + \sum_{j=1}^{n} b_j Z_{ijto}$$

$$a = \text{est } (\alpha_o' + \log P_{xto} + \alpha_1 \log \frac{P_{xto}}{P_{to}} - \alpha_2 \log P_{to})$$

$$a_2 = \text{est } \alpha_2$$

$$b_j = \text{est } \beta_j.$$

To the extent that the variation of the Z_{ijt} are not uncorrelated with Y_{it}, the estimate of α_2 will be affected by the presence of the Z_{ijt} in the equation; therefore it is important to include these variables in the cross section estimation.

In the second stage of pooled estimation, the aggregate relation associating time series market data will be written as

$$(\log X_t - a_2 \log \frac{Y_t}{P_t}) = \alpha_0 + \alpha_1 \log \frac{P_{xt}}{P_t} + e_t$$

This glosses over some complications of aggregation. If the micro relation from the cross section sample is added over i = 1,2, ... N in a straightforward manner, the resulting equation for market aggregates relates the logarithm of geometric mean demand and geometric mean income; whereas the market demand formulation would be between logarithms of arithmetic mean demand and arithmetic mean income. Under some special distributional conditions this discrepancy will not matter in time series estimation. Also, the special variables Z_{ijt} get aggregated over i = 1,2,...N. If they change slowly or not at all in the aggregate, their sums will simply be absorbed in the constant term of the equation. At worst, they should change only on a trend basis and can be included as additive trends in the time-series market equation.

Given the estimate of α_2 (= a_2), a new dependent variable

$$\log X_t - a_2 \log \frac{Y_t}{P_t}$$

is constructed and regressed, in some statistical sense, on log P_{xt}/P_t to get estimates of α_1. This is a standard form of using cross-section together with time series data for estimating both α_1 and α_2. Other procedures that take income distribution and different functional forms into account are available. The main point is that a fruitful sample augmentation can often be found to get round problems of sparseness of samples. Although the example is built round the familiar case of the estimation of demand equations using family budget data, similar procedures can be used for estimation of other econometric relations from cross section data on individual firms, farms, or other units. Too little resort is usually made to the use of cross-section data as a supplement, and this may prove to be decisive in the modeling of developing economies.

Other devices for filling data gaps left in sparse samples are the uses of different forms of a priori information from economic theory, legal rules, institutional relations, engineering information, stable ratio values, and other sources. A case in point is the estimation of a model of Bangladesh. If a plausible model can be estimated from the meager sources for that newly formed nation, it would seem that almost any case of data sparseness could be overcome.

In an unpublished paper, Dr. A.N.M. Azizur Rahman, has estimated and applied (extrapolation, multiplier analysis, general simulation) an equation system of 17 relationships, a not inconsiderable achievement, with results in the range of admissible error that we usually find in applications to developing economies. Data from the accounts of the predecessor state of East Pakistan are used as a base from which to analyze, by model extrapolation, trends and characteristics of contemporary Bangladesh.

What are some of the devices and a priori restrictions used by Dr. Rahman to arrive at a data base suitable for estimation of a (first approximation) tentative Bangladesh Model?

In arriving at a population figure, he used a bench-mark estimate for 1961 and assumed a fixed growth rate to generate an annual series for use in the model. Labor force was estimated from population, assuming that the 1961 participation rate remained fixed. Labor force by sector was distributed annually on the basis of the 1961 distribution, with some external checks from direct observation in the agricultural sector to see if contradictory information were being given by this method. A further simplification was introduced by using labor force instead of employment statistics as measures of factor inputs.

To estimate savings, the whole-Pakistan historical ratio to GNP were used for each annual observation. Given a series on GNP, a series on savings could be estimated.

Published data on rice and wheat availability were used to estimate total food consumption. Capital inflow estimates were estimated from published trade data on current account balance. Export and import data, which were relatively abundant for the whole of Pakistan in the historical period were used for the estimation of Bangladesh trade statistics.

The use of data generated in this fashion, together with other more direct series, implies some degree of systematic variation in many of the relationships estimated, from the process of data creation itself. This is, indeed, a familiar problem in such cases. It is, therefore, not proper to argue on the basis of goodness-of-fit statistics that the model is useful or not useful. It is a matter of validation testing of the performance of the system as a whole in extrapolation and other simulation calculations. Some of the component series are independently observed with moderately good precision, and it is important to see how the system as a whole deals with these in validation tests. This is what Dr. Rahman has done.

A priori information to overcome data sparseness was used in the Bangladesh study by restricting the coefficients in the manufacturing production function to the case of constant returns to scale. This reduced by one, the number of parameters to be estimated and possibly avoided a serious case of collinearity between labor and capital inputs. Additionally, the manufacturing estimates were checked, satisfactorily, against those obtained from a cross-section sample estimate in India. Similarly the parameters of the food consumption equation were checked against those estimated independently for West Pakistan.

In this first approximation, there was no detailed public sector in the model. The equations for the public sector should give scope to the liberal use of institutional information on tax rates and necessary public expenditures. Nevertheless, given the limited approach used and the simplicity of the model, the author found good performance measures for main economic magnitudes - most errors less than ten per cent and many less than five per cent.

It was noted previously that in the case of developing countries there will often be gaps in statistical series. Some observations may simply be missing; some series will be collected infrequently; and some may be completely unavailable. Some of these deficiencies can be made up by arithmetic or purely statistical methods. It is often useful to interpolate (linearly) missing values in order to round out all series to be of the same degree of completeness. Examples of this are abundant in the concentration of the Bangladesh Model.

In the series

$$X_1, X_2, X_3, \ldots X_5, X_6, X_7 \text{ etc.}$$

the value X_4 can be estimated by linear interpolation between X_3 and X_5

$$\text{est. } X_4 = \frac{X_3 + X_5}{2} \quad .$$

A better estimation may be supplied by the movement of a proxy series. If a series Z_t, without gaps, is related to X_t, we may estimate

$$X_t = f(Z_t) \qquad t = 1, 2, 3, \ldots, 5, 6, 7 \text{ etc.}$$

by regression methods. The estimate of X_4 will then be

$$X_4 = f(Z_4) \quad .$$

When national accounts are not available for all periods and all components, related components such as production indexes, price indexes, employment inputs, and other related series may be used as proxies. In some instance, data from a related economy of the same type, in the same geographical region, may be used to construct the proxies.

No matter how serious the data gaps may be within the developing economy, there are usually good and plentiful statistics for trade, covering both exports and imports by value and by volume. These data are likely, also, to be more accurate than are other series. For this reason alone, great emphasis should be placed on the estimation of trade relations for exports and imports by commodity and region. Such relationships may be important in their own right because trade looms large in the economic affairs of most developing countries. Export activity is often a good indicator of overall activity, but extra effort should be put into the estimation of the trade relations because they are likely to have a relatively sound data base.

Just a bit below trade statistics in availability and accuracy are data on primary production, especially food and minerals. Data for the primary sector can be estimated entirely or by proxy methods from basic data on physical volume and prices. The latter may often be inferred from world market quotations.

Admittedly the problems are serious, but with much imagination and effort, they can be overcome to the point of producing results, provided we are realistic in expectations and applications.

2.0 MODEL SPECIFICATIONS

Many of the papers in this volume deal with structural characteristics of economic systems that are peculiar to the case of the developing country. In this section we shall outline the specification for a complete model that will prove to be fairly typical of the developing economy. We shall specify it in a form that is readily adaptable to the typical case, but flexible enough so that it can easily incorporate special features that nearly always turn out to be relevant.

The supply side. It is useful to distinguish between two kinds of output - actual and potential. The latter is the amount that could be produced with available facilities, with the restraint usually being shortages of fixed capital. The former is the output stream that is called forth on the basis of existing conditions, usually final demand conditions in contemporary markets.

These equations, for potential and actual output, are often developed for different sectors. These may be many in number, possibly as many as can be handled in input-output tables available, but often aggregated into conventional classifications. An attractive set of classifications is 1. primary, 2. secondary, 3. tertiary. Primary sectors are agriculture, forestry, fishing, and mining. Secondary sectors are manufacturing and construction. Of course, there are many different lines of manufacturing. Finally, tertiary sectors are trade, personal services, electric power, transport, communications, sanitary, water, and government.

We shall assume

$$X^P_{it} = f_i\ (K_{i,t-1}) + e^K_{it} \qquad i = 1,2,3$$

This states that potential output in primary, tertiary, or secondary sectors are functions of capital available and a random error.

X^P_{it} = potential output of sector i

$K_{i,t-1}$ = beginning of period stock of capital in sector i

e^K_{it} = random error.

The potential output function could be linear, log linear, or some curvilinear S-shaped function. These are leading possibilities. The log linear form would be

$$\log X^P_{it} = \alpha_0 + \alpha_1 \log K_{i,t-1} + e^K_{it}$$

To obtain statistics for $K_{i,t-1}$ it will be necessary to cumulate net investment from a fixed base

$$K_{it} = K_{io} + \sum_{j=1}^{t} (I_{ij} - D_{ij})$$

where I_{ij} = real gross capital outlays by sector i in period j

D_{ij} = capital consumption by sector i in period j.

If capital decays geometrically, we would have

$$D_{it} = \lambda K_{i,t-1}$$

An identity associating K_{it} and I_{it} can be derived as follows

$$K_{it}-K_{i,t-1} = I_{it} - D_{it}$$

This is obtained by differencing the expression for K_{it}. It becomes

$$K_{it} - K_{i,t-1} = I_{it} - \lambda K_{i,t-1}$$

$$K_{it} = I_{it} + (1-\lambda) K_{i,t-1}$$

This gives a recursive relation for building up K_{it} on the basis of knowledge of the previous values and I_{it}. One must also know, approximately, the decay rate of capital, λ. A parameter such as this may be estimated from engineering information or general knowledge about the length of life of capital. Such information may transcend country boundaries. Knowledge of the initial value, K_{i0}, is not essential in linear systems, for it gets combined with the constant term in an equation. But if the system is log linear or otherwise curvilinear, it is essential to have an initial wealth estimate for the economy and its separate departments.

Since labor is often in abundant supply in the developing country and capital is the main limiting factor, we may split the production process into two parts - one being the transformation of capital into output and the other being the labor requirements per unit of capital, in other words the necessary labor to man the capital.

$$L_{it} = \mu K_{i,t-1} + e^{L}_{it}$$

L_{it} = labor force attached to sector i

e^{L}_{it} = random error.

This function, of course, could be non linear.

For the agricultural sector, and also possibly mining, some more distinctive specification forms may be justified. It is not altogether ridiculous to make the agricultural potential output function dependent on equipment and acreage, as limiting capital factors, but it is undoubtedly better to model agricultural production directly in terms of natural factors. Also, if the agricultural products being considered gare widely traded international commodities, supply reactions to world prices would also be in order. This would apply equally well to mineral products.

The agricultural case may be split into two levels - acreage and yield/per acre. The acreage decision can be expressed as

$$A_t = f\,(PW_t,\ PW_{t-1},\ PW_{t-2},\ \ldots q_t, q_{t-1}, \ldots) + e^{A}_{t}$$

A_t = acreage

PW_t = world price of the commodities in question

q_t = price of supply substitutes. e^{A}_{t} = random error

The lag distribution is likely to be long - five years or more - in the case of acreages for commodities that have a long gestation period. This would be characteristic of tree crops (coffee, cocoa, rubber). It would also be the case for mineral products, where there is a long lead time in mine preparation.

For field crops and poultry, the lag is much shorter, probably not more than one year in physical terms. For these, a short lag distribution of no more than four years would seem to be in order.

The simplest way to introduce lag effects is through the use of the so-called Koyck transformation.

$$A_t = \alpha + \alpha_0 \, PW_t + \beta_0 \, q_t + \lambda \, A_{t-1} + e_t^A$$

This is, apart from error cumulation, also expressible as

$$A_t = \frac{\alpha}{1-\lambda} + \alpha_0 \sum_{i=0}^{\infty} \lambda^i \, PW_{t-i} + \beta_0 \sum_{i-0}^{\infty} \lambda^i q_{t-1}$$

More general lag distributions could also be used. The main point is to allow for the possibility of a fairly long lag response to world prices.

Yield per acre depends on current price, to show the degree of favorableness of marketing now, and on natural factors. The latter may be rainfall, storms, temperatures, and pestilence. We write this as

$$Y_t = \gamma_0 + \gamma_1 \, PW_t + \gamma_2 \, N_t + e_t^Y$$

Y_t = yield

PW_t = world price of the commodity

N_t = natural conditions (good weather or typhoons, or other).

An identity produces an estimate of the actual goods flow in sector 1.

$$S_t = (Y_t)\ A_t$$

S_t = supply of the commodity

Y_t = yield A_t = acreage Y_tA_t = current flow of goods.

After potential output equations are estimated from observational data taken from periods of "peak" level activity, we need to specify equations of actual output. These are expressed as functions of end-use, final demand. In some sense, these equations stand for inventory processes, which are hard to model for developing countries because of lack of good data on stocks or stock changes. The equations show the relation between production and final consumption demand. That is the sense in which they reflect inventory movements.

The final demand equations can be written as

$$X_{it} = f_i\ (X_t - E_t,\ E_t,\ Z_t) + e^x_{it}$$

X_{it} = real output originating in sector i (shipments or demand)

$X_t - E_t$ = real domestic expenditures

E_t = real exports

Z_t = exogenous demand variables

The explanatory variables can be specialized in order to conform to demand situations in sector i. The general variable $X_t - E_t$ consists of real domestic expenditures on consumption, investment, and public sector purchases. These are components of GNP. For agricultural output, the appropriate demand component would be principally consumer expenditures. The shipment abroad of agricultural raw materials would be indicated by the presence of E_t (or a subset of E_t) in the equation for agricultural output.

Since X_{3t} includes much infrastructure activity (electric, gas, water, sanitary, communications, transport), exogenous variables representing public action in support of the infrastructure will be important in the composition of Z_t for the relevant equation. Also, tertiary activity may affect secondary sector activity and vice versa. It is possible to go further into the details of inter relatedness of the different production sectors, at least as far as interpretation is concerned. In a general input-output framework, we have

$$(I-A)\ X_t = F_t$$

where

X_t = gross output vector

F_t = final demand vector

A = input-output matrix

I = identity matrix.

This is the standard linear equation system of input-output analysis. By inversion, we obtain

$$X_t = (I-A)^{-1} F_t$$

The final demand equations detailed above can actually be interpreted as explicit extensions of these input-output transformations, the main differences being that some of the exogenous variables in Z_t may not be direct components of F_t and that X_t represents gross output in the input-output system, while X_{it} is interpreted as real value-added in sector i in the set of final demand equations. This latter differences can be reconciled easily if ratios of gross output to value added are approximately constant.

The demand side

The components of GNP of $(X_t - E_t)$ (consumer, investment and government expenditures) plus exports. As in the case of models of industrial market economies, there will be consumption, investment, export, and import functions in the developing country model. Although more attention may be paid to the supply side in the developing country models, the demand side is important and cannot be neglected. The distinctions are partly semantic, however, because demand for capital goods (investment) and some imports (capital, fuel, materials) are important in determining supply of goods through the production process.

The aggregate consumption function can be specified in the usual way as

$$C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 C_{t-1} + e_t^c$$

C_t = real consumer expenditures

Y_t = real disposable income

The coefficients are denoted as general values α_i, and have no particular relation to the same notation used in specification of general supply side equations.

It may not be easy or possible to obtain satisfactory data for disposable income. In such an event, Y_t would have to be chosen as the conceptually nearest aggregate that is available. The presence of lagged consumption in this equation indicates that consumer spending is equivalently a distributed lag in past income values. In order to come closer to the special characteristics of the developing economy, it will be useful, if possible, to break Y_t into two components

$$P_t Y_t = W_t + P_t$$

disposable wage income and disposable non wage income.

If W_t and P_t, separately deflated and adjusted for taxes and transfers if possible, were introduced as two independent explanatory variables in the consumption function, we would be taking account of an aspect of income distribution. This is of great significance for the developing economy and should be introduced because of the tendency of income inequality to be more pronounced than in the case of industrial market economies. The separate income variables W_t and P_t indicate the effects of the factorial share distribution of income, distinct from the size distribution. Although the latter aspect is more important and more sensitive, size distribution statistics are not readily available on a repeated basis, therefore factorial shares are used, since they are more likely to be available. In case W_t and P_t are strongly collinear, as often happens, a priori estimates of the relative marginal propensities to consume might be used. Instead of two separate variables, we would use the combined variable.

$$W_t + \frac{\alpha_{1p}}{\alpha_{1w}} P_t$$

where α_{1p}/α_{1w} is the ratio of marginal propensities to consume out of the two income types. Family budget data, in the form of Engel curve analysis, may be used to estimate α_{1p}/α_{1w} .

A third distinction in factorial income shares is also important, namely, the rural or agricultural share. A threefold decomposition of income

$$Y_t = W_t + P_t + A_t^*$$

is even more revealing. In this case, W_t and P_t refer to the non agricultural sector, while A_t^* stands for total agricultural income. It is usually the case that the marginal propensity to spend out of A_t^* is the lowest of all three.

An important component of spending in the developing country is in the form of collective, public consumption. We shall understand the previous equation to refer only to the private sector. By contrast $C\ G_t$ will stand for public consumption, or government purchases of current goods and services. These are usually large. A simple hypothesis for the explanation of $C\ G_t$ is that it is limited by public revenues, mainly taxes.

$$C\,G_t/N_t = \beta_0 + \beta_1\, T_t/P_t N_t + e_t^{CG}$$

This is expressed on a per capita basis because public spending must be related to the population base, on whose behalf the spending is for.

Fixed capital formation, like consumption, will be divided into a private and public part. The latter is important in accounting for infrastructure investment. Some form of stock-adjustment or flexible accelerator investment function is likely to be the most relevant, extended, perhaps for the flow of imported capital.

$$I_t = \gamma_0 + \gamma_1\, X_t + \gamma_2\, K_{t-1} + \gamma_3\, K\,F_t + e_t^I$$

I_t = real fixed capital expenditures (gross)

X_t = real GNP

K_{t-1} = stock of fixed capital[1)]

$K\,F_t$ = real financial capital inflows from abroad

Investment in particular sectors will be needed, if capital series for production functions are to be separately generated. In such cases, X and K would be suitably defined by sector to match the sector composition of I.

Finally, domestic spending relations will be rounded out by consideration of public investment.

---

1) See the identities above, relating K_t and I_t, in accounting relations.

$$I\,G_t = \delta_0 + \delta_1\,(T_t/P_t) + \delta_2\,K\,F_t + \delta_3\,N_t + e_t^{CG}$$

In this relation, we have suggested that public investment depends on real revenues (T_t/P_t), real capital inflows from abroad $(K\,F_t)$, and population size (N_t).

The demand or expenditure side is completed with an analysis of foreign trade. We need export and import equations. Exports, to the extent that they consist of primary products sold in major world markets dominated by industrial countries will depend on general activity variables abroad. If the export is highly specified (rubber, tin, coffee, cotton, etc.) it will depend on some specific line of activity abroad; e.g. cotton exports will depend on textile production in the major importing countries. Price will also be a factor, but many of these goods are inelastic in demand. Inflation effects may be taken care of by using real price, i.e. own price divided by a general price index abroad, or a closely related price. The related price is likely to be substitute commodity such as tea for coffee, synthetic fibers for cotton, aluminium for copper, etc.

A general export equation is therefore

$$E_t = \zeta_0 + \zeta_i\,(XW)_t + \zeta_2\,\left(\frac{r\,p_e}{pw}\right)_t + e_t^E$$

E = real export

XW = world industrial production (in importing countries)

r = exchange rate

P_e = export price

PW = world general price level.

There should be a group of export equations to cover the different types of goods exported, but they will have the same general characteristics. In most cases, the relatively small developing country is likely to be a price taker. Recently, the oil exporting countries have become price makers, but this is unusual.

Imports are likely to have more specific structural characteristics. They may be associated with import substitution. They may involve basic food. Capital imports will depend on reserve position and foreign capital transfers. The usual function will be

$$M_t = \eta_0 + \eta_1 X_t + \eta_2 \left(\frac{pm}{rp}\right)_t + e_t^M$$

M = real import

X_t = real domestic product

P_m = import price

r = exchange rate

p = general domestic price

If import substitution is prevalant or policy stimulated, there may be an inverse relation between M and X. In other words, we would have

$$\eta_1 < 0_1$$

X_t would be more strictly defined for some import lines. It may be total domestic food production (also with a negative coefficient), total domestic consumption, total domestic capital formation, etc.

If M consists strongly of capital goods, an exogenous variable KF_t should be added. This stands for capital flow (for fixed capital formation) from abroad.

Another variable for the import equation is "capacity to import". It is usually evaluated as the value of exports divided by import price. This shows the real import purchasing power of exports.

$$\frac{(Pe\ E)_t}{(Pm)_t} .$$

<u>Wages, prices and income:</u> For the production of goods, factors are required and paid incomes. The incomes received are used to purchase the output according to the categories outlined in the previous section. From the supply side relationships we determine labor requirements via the production functions. The implied employment levels together with market wage rates make up the components of wage income.

In the developed economy, wages are viewed as the outcome of a bargaining process, pitting employer against employee.
This bargain is sensitive to unemployment, which represents the discrepancy between supply and demand. The money wage rate, also follows movements in prices to provide "cost-of-living" catch-up, particularly in inflationary environments. This kind of relationship seems to be less appropriate for the developing country. We, instead, assume that wages will conform more to long run trend prices. Real wage rates should follow productivity

$$\Delta \log \left(\frac{W}{P}\right)_t = \sim_0 + \sim_1 \Delta \log (X/L)_t + e_t^W$$

w/p = real wage rate

X/L = productivity (output per man)

This is an economy-wide relationship and could be split into different patterns for individual sectors. By forming first differences in logarithms, we approximate percentage changes. This is a noninflationary pattern. It states simply that workers improve their standards by participating in productivity gains. From the estimation of w_t here and L_t from the production relations on the supply side, we can form wage income as

$$W_t = w_t L_t$$

This is an accounting identity.

It remains to determine the price level for the developing economy case. Just as the specification of wage rate determination departs, in this case, from the typical short run labor market bargaining analysis of the advanced industrial economy, so does the determination of price level. The prevalent theory of price determination in the industrial economy is the close calculation of mark-up over unit labor cost or unit total cost, both long and short run, with allowances for response to cyclical fluctuations in capacity pressures and import costs. An alternative theory is provided by the classical quantity theory of money which makes prices proportional to money supply along a steady full-employment growth path. This theory is singularly inappropriate in short run analysis, except possibly when monetary expansion gets out of control and the speed of the printing presses dominates everything else.

An eclectic theory of price determination for the developing country takes parts of these different specifications as appropriate to the case at hand. Changes in price level for the developing country may be assumed to respond to changes in velocity of money circulation (a quantity theory aspect), to the pressure of domestic demand on productive facilities (short run capacity utilization) and to import prices. The latter will be denominated in local currency; therefore movements of exchange rates are reflected in domestic prices through the conversion of world prices to domestic currency units. An equation representing these eclectic ideas is

$$\Delta \log p_t = \Theta_0 + \Theta_1 \Delta \left(\frac{pX}{MS}\right)_t + \Theta_2 (X/X^p)_t + \Theta_3 (p_m/r)_t + e_t^p$$

p_t = general price level

MS_t = money supply

X_t = aggregate real output

X_t^p = aggregate real potential output

pm_t = input price level measured in foreign currency units

r_t = exchange rate

In different countries and in different situations, these reaction coefficients Θ_1, Θ_2, Θ_3 may vary, showing importance of one form of price influence compared with another.

The price of exports will follow general domestic price movements, but if the export item is widely traded on world markets, as is the case with many primary products, it will follow world price developments.

$$p_{et} = \Theta_0^e + \Theta_1^e p_t + \Theta_2^e (PW)_t + e_t^{pe}$$

Apart from the development of a monetary sector, these equations complete the behavioral and technological specifications of the typical LDC model. There will be equations for taxes, transfers, and accounting identities to close the system.

Taxes, and transfers as well, they will depend on nominal income flows. If they are personal taxes, they will depend on personal income. If they are indirect taxes they will depend on output value. In any event, these equations should not be simply fit to the data to obtain statistical relationships of the usual sort. They should be interpreted as far as possible directly in terms of statutory regulations with changes frequently introduced as policy changes. A typical relationship would be

$$T_t = {}_0 + {}_1 (P_t Y_t + T_t) + e_t^T$$

This equation is in nominal terms, whereas the behavioral equations are largely in real terms.

T_t = taxes

$P_t Y_t$ = nominal value of disposable income

The main identities will be

definition of GNP $X_t = C_t + I_t + IG_t + CG_t + E_t - M_t$

(total expenditure)

aggregate production $X_t = X_{it} + X_{2t} + X_{3t}$

The dependent or endogenous variables generated by this system are

X^p_{it}	$i = 1,2,3,$	potential output
X_{it}	$i = 1,2,3,$	actual output
K_{it}	$i = 1,2,3,$	capital stock
I_{it}	$i = 1,2,3,$	investment
D_{it}	$i = 1,2,3,$	capital consumption
L_{it}	$i = 1,2,3,$	Employment
A_t		Acreage
Y_t		yield
S_t		supply of crops
C_t		private consumer expenditure
W_t		wage income
P_t		nonwage income
Y_t		real disposable income
CG		public consumer expenditures
IG		public investment outlays
E_t		real exports
M_t		real imports
w_t		wage rate
p_t		price level
p_{et}		export price level
T_t		tax revenues (less transfers)

This is a general prototype system. Money supply is treated as though it is exogenous, but it could be modeled in terms of the internal deficit, the external deficit, and monetary policy. The exchange rate is considered to be exogenous, and the domestic interest rate is not explicitly introduced. In any realistic case, there will be special features, additional disaggregation and other variables. This system, however, serves as the prototype for the case of developing countries.

3.0. PARAMETER ESTIMATION AND HYPOTHESIS TESTING

Economic theory, as pointed out in section 2, provides us with suggestions for possible relationships among the variables of a model. Econometrics is concerned with confronting these suggestions with actual data, with testing the validity of these relations, and with estimating the parameters involved. The most important statistical tool employed in this stage of model building is regression analysis. In this section we shall take up the underlying ideas of regression methods to understand their potentials and limitations in the context of model building for developing economies.

As a starting point for understanding the basic concepts of econometric estimation we analyze the fundamental consumption - income relationship which states that private consumption expenditures are determined to a great extent by the amount of disposable income available. Employing data for Mexico we plot what is called a scatter diagram which, in fact, shows a pronounced systematic relationship between real private consumption expenditures and real disposable income. This systematic component can be approximated by a linear equation.

$$(3.1) \qquad Y_t = \alpha + \beta X_t$$

where α and β are parameters determining the position of the equation, Y_t the variable symbol for real private consumption at time t, and X_t the variable symbol for real disposable income at time t. In our example we use pairs of observations on Y and X, later referred to as sample, between 1950 and 1973, the sample period. Inspecting the scatter diagram (Fig. 1) we observe that

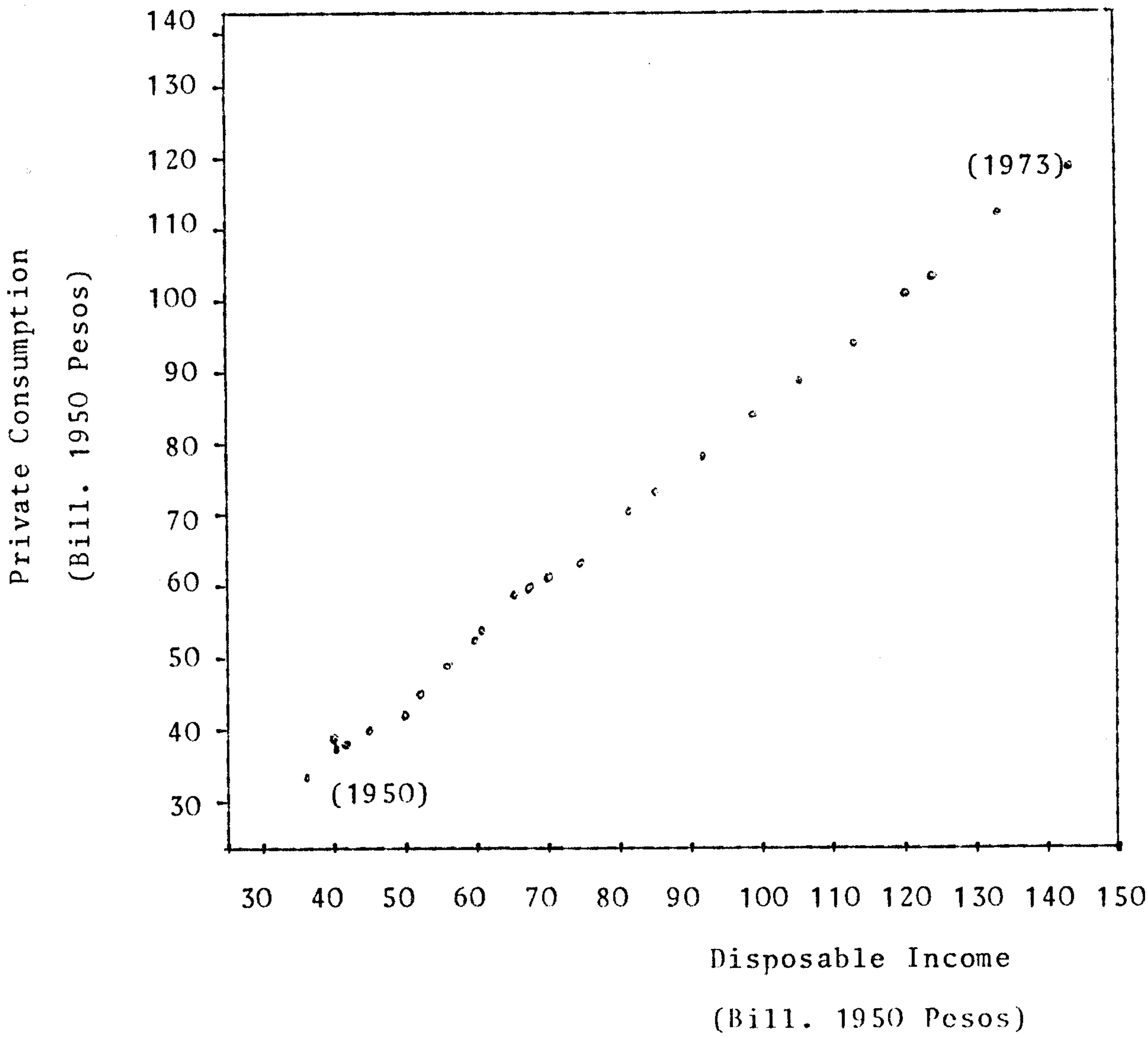
Private Consumption
(Bill. 1950 Pesos)
140
130
120
110
100
90
80
70
60
50
40
30
(1973)
(1950)
30 40 50 60 70 80 90 100 110 120 130 140 150
Disposable Income
(Bill. 1950 Pesos)

Fig. 1

the specified linear relation (3.1) does not hold strictely, but is more or less disturbed. We, therefore, extend the specified relation by adding to the systematic component what is called a disturbance term or nonobserved random error e_t taking care of the discrepancy between actual observations and the systematic component. The rationale for this random component are ommitted or unobservable relevant factors, changes in human behavior or measurement errors in the data. Because of the presence of this error term we are dealing now with a stochastic relation which is called linear regression model and in its simplest form is described by

$$(3.2) \qquad Y_t = \alpha + \beta X_t + e_t \quad .$$

Observations on Y and X can be made over time, yielding time series data, or over individuals, yielding cross section data, or can be a mixture of both types of data.

So far the parameters α and β in our stochastic relation are unknown and we shall try to determine their values by the information provided by the sample. This process is called parameter estimation.

A pragmatic suggestion for the determination of the unknown parameters would be to choose those values for α and β which minimize the discrepancy between observed values and those computed by the systematic component of the regression equation. If we employ as a measure of discrepancy the sum of the squared errors, we minimize

$$(3.3) \qquad \sum_{t=1}^{T} (Y_t - \alpha - \beta X_t)^2$$

with respect to α and β and get as estimators that are formulas to calculate the unknown parameters from the sample values,

(3.4a) $$\hat{\beta} = \frac{\Sigma(X_t-\bar{X})(Y_t-\bar{Y})}{\Sigma(X_t-\bar{X})^2}$$

(3.4b) $$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

The "hat" on α and β indicates that these are estimators for the unknown parameters α and β. According to our minimization criterion these are least squares estimators.

It turns out that this pragmatic suggestion for estimation of the unknown parameters has highly desirable statistical properties. If we assume that the error terms are independent and identically distributed with mean zero and variance σ^2, then the above estimators have the so called BLUE (best linear unbiased estimators) property. Unbiasedness means that the exptected value of those estimators is equal to the true parameter values and the property best states that no other unbiased estimators have a smaller variance than these least squares estimators. The sample variances for the estimators above are given by

(3.5a) $$\text{var}(\hat{\beta}) = \frac{\sigma^2}{\Sigma(X_t-\bar{X})^2}$$

(3.5b) $$\text{var}(\hat{\alpha}) = \sigma^2\left[\frac{1}{T} + \frac{\bar{X}^2}{\Sigma(X_t-\bar{X})^2}\right]$$

As the variance of the error term σ^2 is an unknown parameter too, an unbiased estimator is given by

(3.6) $$s^2 = \frac{1}{T-2}\,\Sigma(Y_t - \hat{\alpha}-\hat{\beta}X_t)^2$$

Replacing, in (3.5a) and (3.5b), σ^2 by s^2 we get estimators for the sample variances of $\hat{\alpha}$ and $\hat{\beta}$, $s^2_{\hat{\alpha}}$ and $s^2_{\hat{\beta}}$, respectively.

If we add the assumption that the error terms are normally distributed we can make probability statements about the estimated parameters, as the ratios

(3.4 $$\frac{\hat{\beta} - \beta}{s_{\hat{\beta}}} \quad , \quad \frac{\hat{\alpha} - \alpha}{s_{\hat{\alpha}}}$$

have a t-distribution with (n-2) degrees of freedom. This enables us to construct confidence intervals for the true values and to test alternative hypotheses about the true parameter values.

A measure commonly used to describe how well the sample regression line fits the observed data is the coefficient of determination, denoted by R^2 and defined as

$$R^2 = 1 - \frac{\Sigma(Y_t - \hat{\alpha} - \hat{\beta}X_t)^2}{\Sigma(Y_t - \bar{Y}_t)^2}$$

R^2 cannot be negative or greater than one. A zero value indicates the poorest and a unit value the best fit that can be attained.

An important assumption in the linear regression model as to the error term is the absence of correlation between different errors. A widely used test for the presence of autocorrelation of first order was developed by Durbin and Watson, and is defined by

$$DW = \frac{\sum_{t=2}^{T}(\hat{e}_t - \hat{e}_{t-1})^2}{\sum_{t=1}^{T}\hat{e}_t^2}$$

where $\hat{e}_t$ denotes the computed errors from the estimated regression.

A rough idea about the use of the DW-statistic is given by the fact that values around 2 indicate the absence of autocorrelation; whereas values near 0 and 4 indicate positive and negative autocorrelation, respectively.
Although this exposition of the linear regression model was restricted to two variables there is no difficulty in extend in this analysis to more variables. Mention should be made of the fact that for a system of interdependent equations, so-called simultaneous estimation methods are available, which in most cases consist of multiple application of least squares estimation.

Now we apply the foregoing analysis to the consumption-income relationship for Mexico. We denote by CP private consumption expenditures at constant prices and by YD/PGNP disposable income at current prices divided by the GNP-deflator. The linear regression model (3.2) applied to these variables with annual observations between 1950 and 1973 yields the following results:

$$CP = \underset{(.5600)}{5.75131} + \underset{(.0067)}{.80147}\ YD/PGNP$$

$R^2 = .998$ S.E.=1.05925 DW=1.19

In order to maintain accuracy in competition, several digits, many of them to thright of the decimal print, are carried through to the final stage of calculation. Actually, we cannot be very sure of many of these digits. At fact we can estimate

$$\hat{\alpha} = 5.75$$

$$\hat{\beta} = 0.80$$

$$S.E. = 1.06$$

The listing of several digits in the final competition is not meant to convoy an impression of the achievement of unusually high precision.

The estimates for the unknown parameters α and β are 5.75131 and .80147. The number in brackets under the parameter values are the estimated sample standard deviations of the estimates from the true parameter values. The coefficient of determinations is .998 which means that 99.8 % of the variance of the private consumption expenditures time series is explained by the above regression model. S.E. denotes the estimated standard deviation of the error term, which in this equation is 1.05925 billion Pesos at constant prices. The value of the Durbin-Watson statistic of 1.19 indicates the presence of positive autocorrelation in the residuals and might indicate the need for a better model specification. The estimated parameters are usually tested to determine, if for a given level of significance, there is enough evidence to indicate that the true parameters are different from zero. A null hypothesis is formulated, stating that the true parameter value is zero and an alternative hypothesis that it is different from zero. According to (3.4) the test statistic for β is

$$\frac{|.80147 - 0.|}{.0067} = 119.52$$

and the test statistic for α is

$$\frac{|5.7531 - 0.|}{.5600} = 9.74 \quad .$$

These have t-distributions with (24-2) degrees of freedom. The tables for the t-distribution show that for a level of significance of 95 % and 22 degrees of freedom the null hypothesis is rejected when the value of the test statistic is larger than 2.074. This is the case for both estimated parameters which leads us to the conclusion that the true parameters are significantly different from zero. This means that there is strong statistical evidence that disposable income has an influence on private consumption expenditures.

Besides statistical tests on the estimated regression equation, we check, in addition, the parameter values for their plausibility from the point of view of economic theory. The coefficient is the marginal propensity to consume (MPC), indicating the effect of a unit increase of disposable income on consumption expenditure. Its estimated value agrees with what we would consider to be reasonable, from the view point of international comparisons .

There are some indications that the above specified consumption - income relationship can be improved. One indication comes from economic theory, which suggests that decisions about current consumption expenditures are not only based on current but also on past levels of income. The other is the evidence of autocorrelation in the residuals, which often hints at the existence of an omitted variable. Therefore we refine the first specification of the consumption - income relationship by adding one-period lagged consumption as an explanatory variable. We know from section two that this is a way to estimate a geometric lag distribution. This means that current consumption is determined by current and past disposable income, with geometrically declining weights. We get the following regression result:

$$CP = .20067\ CP(1) + .64984\ YD/PGNP + 4.65384$$
$$\quad (.1123) \qquad\qquad (.0851) \qquad\qquad (.8139)$$

$R^2=.999$ DW=1.88 S.E.=1.01012

The sample size is reduced by one in order to include CP(1). The statistical properties of this specification are favorable except for the fact that the first coefficient is on the verge of statistical insignificance. As the economic content is more satisfying than in the first relationship, we put this equation into the model proposed in section four.

This is a typical example of the process of econometric model building. We start with a hypothesis of economic theory, confront this hypothesis with empirical evidence, which enables us to maintain or reject the proposed relationship. The experience we gain during this process leads us to reformulate our theoretical hypothesis and to continue this feedback procedure between economic theory, data, and statistical method.

The general availability of modern computer and special econometric software packages makes running a regression almost as easy as making some simple algebraic operations on a desk calculator.

There is some danger of misuse of regression analysis if we employ it only in a mechanical way. A word of caution seems to be appropriate, to be aware of limits and dangers when working with this powerful instrument.

Strictly speaking the regression model is valid only if the following assumptions are fullfilled:

1. Correct specification of the error process.
2. Observations free of measurement error.
3. No prior information available about the parameters.
4. Constant parameters over time.

Because of the nature of the subject we are dealing with, none of these assumptions in general is completely valid in an econometric model. This does not completely bring into question the application of the proposed statistical methods., if we are aware of the deficiencies involved and if we have an idea of the extent to which they are violated.

There are a number of methods available to deal with more complicated error processes, if we have some evidence of the kind of error structure involved. The same holds for measurement errors, if we have some information about the precision of the data used. If there is prior information available about the parameters to be estimated, we are well advised to include this information explicitly in the estimation process as it helps us to regain some of the precision we loose by data deficiencies.

Most questionable is the validity of the assumption of time-invariant parameters when we deal with economic systems. Attitudes shift: technology changes; and political factors tend to alter the status quo. We want to stress the importance of this problem by investigating the stability of the estimated parameters in the equation for private consumption expenditures of the model for Mexico.

The following experiment is designed: First, we fix the starting time of the sample observations with 1950 and vary the end time from 1965 to 1973. This case is referred to as an expanding sample regression. Second, we fix the length of the sample with 16 observations and move the starting time of the sample from 1950 to 1958. This case is referred to as a moving window regression. Table 1 and Table 2 show the results of both approaches. In both cases we observe a considerable variation of the estimates due to variations in the sample period. These variations become especially evident in the moving window case, as they are smoothed out in the expanding sample version. The forecasting error is considerably lower when a moving window sample is used.

The above example is not atypical for many relations used in econometric models. It should not discourage the user of regression methods in parameter estimation but make him aware of the problems and stimulate similar investigations to get an idea about the reliability of the estimated models. A considerable amount of research is under way to develop parameter estimation techniques which try to take time variation of the model parameters explicitely into account. Together with research efforts on estimation methods for deficient data, new estimation methods should be employed which make optimal use of all information available in the process of parameter estimation. In contrast to the widespread opinion that a poor data base allows only for simple estimation methods, it is suggested that sophisticated techniques are required.

Tab. 1: Expanding sample regression for the consumption equation for Mexico:
CP = A1 CP (1) + A2 YD/PGNP + A3

Sample	Coefficient			S.E.	DW	R2	R2C
	A1	A2	A3				
50-65	.184 (1.35)	.687 (6.22)	3.481 (3.16)	1.044	1.78	.994	.994
50-66	.209 (1.62)	.661 (6.45)	3.700 (3.56)	1.025	1.79	.996	.995
50-67	.220 (1.79)	.649 (6.78)	3.811 (3.90)	.997	1.81	.996	.996
50-68	.242 (1.99)	.625 (6.67)	4.059 (4.25)	1.002	1.81	.997	.997
50-69	.265 (2.13)	.601 (6.33)	4.350 (4.52	1.032	1.71	.997	.997
50-70	.255 (2.16)	.610 (6.83)	4.311 (4.63)	1.007	1.81	.998	.998
50-71	.222 (1.97)	.637 (7.32)	4.697 (5.61)	1.004	1.72	.998	.998
50-72	.200 (1.75)	.652 (7.54)	4.541 (5.36)	1.026	1.74	.998	.998
50-73	.201 (1.79)	.650 (7.64)	4.654 (5.72)	1.010	1.88	.999	.999

t-ratios in ( )

R2C is R2 adjusted for degrees of freedom.

Tab. 2: Moving window regression for consumption equation for Mexico
CP = A1 CP(1) + A2 YD/PGNP + A3

Sample	Coefficient			S.E.	DW	R2	R2C
	A1	A2	A3				
50-65	.184 (1.35)	.687 (6.22)	3.481 (3.16)	1.043	1.78	.994	.994
51-66	.192 (1.33)	.672 (5.97)	3.893 (3.10)	1.060	1.38	.995	.994
52-67	.269 (2.22)	.626 (6.75)	2.826 (2.52)	.879	1.62	.997	.997
53-68	.317 (2.40)	.573 (5.71)	3.303 (2.88)	.925	1.76	.997	.997
54-69	.366 (2.61)	.522 (4.89)	3.957 (3.33)	.971	1.86	.997	.997
55-70	.322 (2.25)	.557 (5.28)	4.240 (3.19)	.969	1.88	.998	.997
56-71	.201 (1.63)	.635 (6.86)	5.948 (5.18)	.885	1.90	.998	.998
57-72	.078 (.65)	.728 (8.22)	6.961 (5.85)	.819	1.74	.999	.998
58-73	.017 (.13)	.770 (8.07)	7.723 (5.82)	.790	1.78	.999	.999

t-ratios in ( )

4.0 STEPS IN ECONOMETRIC MODEL BUILDING

In this chapter we shall consider the composition of economic relationships in a model. It will be seen that the same iterative process between economic theory, data base, and statistical methodology which led us to accept a single relationship will be used to establish a system of equations which constitute a model.

Models are, by definition, approximations of reality. A criterion for their usefulness, if they are acceptable approximations, is determined by the purpose for which the model is to be used.

Step one in the model building process is the specification of the scope of the model. This involves decisions about the degree of disaggregation of economic activity in the model and the inclusion or exclusion of specific sectors. The scope of a model depends on the requirements for use of the model. The proposed scope of a model determines the data base needs. Unavailability of certain data or their limited reliability may lead to a reformulation of the scope of the model. The main source for the data base of macro-econometric models is the national income accounts. Additional information for the data base is provided by special statistical publications about the price, financial, trade, and employment sectors.

Step two involves the distinction between input and output variables of a model. In econometrics the former are referred to as endogenous variables and the latter as exogenous variables.

This distinction again depends on the purpose and environment of the model. Exports, for example, may be considered exogenous in the framework of a national model but are obviously endogenous in the scope of a world trade model. Other variables in a model such as public expenditures or tax rates may be considered exogenous, because they may be controlled by policy makers.

Step three is concerned with formulating a mathematical equation for each endogenous variable. Two types of equations are encountered in an econometric model: stochastic equations of the type described in the previous section. These are also called behavioral equations, as they are usually given a causal interpretation for the behavior of certain economic agents. Deterministic equations or identities follow from definitional relationships in the data base.

Step four deals with estimation of the unknown parameters in the stochastic equations and tests for their significance, based on the evidence of the sample information. This step was extensively described in the previous section.

To illustrate the steps of the model building process described so far, we present a simple example for a macro-econometric model. We assume the following scope and data base for our model:

C	Consumption
DEP	Depreciation
GNP	Gross national product
I	Investment
K	Capital stock

M	Imports
TI	Indirect taxes
TD	Direct taxes
X	Exports
Y	National income
YD	Disposable income

Next, we decide to consider exports as exogenous variables. They are mainly determined by international economic activity. Similarly, indirect and direct taxes are exogenous, as they are subject to the control of policy makers, as is depreciation. We therefore need equations for the remaining endogenous variables, namely consumption, gross national product, investment, capital stock, imports, national income, and disposable income.

We start with the main definitional identity of the data base, which states that GNP is determined by the sum of consumption, investment, and net foreign demand, that is exports minus imports. We then specify stochastic behavioral equations to describe the level of consumption and investment expenditures and import requirements. As exports are exogenous, we use this main identity to calculate GNP.

Consumption expenditures are specified to be determined by a lag distribution in disposable income. Investment activity depends on maintenance requirements of the existing capital stock and its adjustments to output changes. Imports are assumed to be approximately proportional to domestic economic activity. The α, β and γ parameters in these three stochastic relations are estimated by the regression methods described in the previous section and tested for significance.

The remaining three endogenous variables are determined by the definitional identities which hold for the data base. The complete model is described by the following equation system:

$$C = \alpha_o + \alpha_1 \; C(1) + \alpha_2 \; YD + e_1$$

$$I = \beta_o + \beta_1 \; (GNP-GNP(1)) + \beta_2 \; K(1) + e_2$$

$$M = \gamma_o + \gamma_1 \; GNP + e_3$$

$$GNP = C + I + X - M$$

$$Y = GNP - DEP - TI$$

$$YD = Y - TD$$

$$K = (1 - \delta) \; K(1) + I$$

<u>In step five</u> of the model building process we turn to the <u>solution</u> of our specified model, which we understand calculation of the values of the endogenous variables of a model for given values for all remaining variables of the equation system. These variables encompass not only the exogenous variables but also lagged endogenous variables, as can be seen from our example. If an equation system is linear in all variables, then an explicit representation of the endogenous variables as functions of lagged endogenous and exogenous variables can be given. This is the so called reduced form of a model. Typical econometric models, however, are non-linear in the variables. Such a system is solved by the Gauss-Seidel method, an iterative solution technique, which starts which initial values for the endogenous variables to calculate a first solution of the system, which in turn

is used in the next step for a modified solution. This procedure is continued until some convergence criterion is fulfilled.

If we are able to produce a solution of a model we can proceed to step six, the validation of a model. In this step we compare the model with the behavior of the real system the model was designed for. Two procedures are applied to test the validity of a system. One is concerned with pseudo-predictions, or ex post forecasts, where we use the model to generate forecasts in the sample period. We compare the actual paths of the endogenous variables with the path generated by the model to get an idea about the degree of historical verification. Commonly used measures to compare the computed values $\hat{x}_t$ with actual values x_t are the mean square error,

$$MSE = \frac{1}{T} \Sigma (x_t - \hat{x}_t)^2$$

the root mean square error,

$$RMSE = \sqrt{\frac{1}{T} \Sigma (x_t - \hat{x}_t)^2}$$

or the mean absolute percentage error

$$MAPE = \frac{1}{T} \Sigma \frac{| x_t - \hat{x}_t |}{x_t} .$$

The other validation procedure involves what is called multiplier analysis. An arbitrary solution path of the model is generated, the control solution. Then one or several exogenous variables are changed, and a new solution path is calculated. The difference between control solution and the new solution is called multiplier, as it describes the effect of changes in exogenous variables on the endogenous variables. This multiplier results are checked for their economic plausibility.

Both validation procedures, ex post forecasting performance and multiplier results,are used to judge the quality of the complete model. Implausible results may provide an incentive for a reformulation of the model starting again with step one.

Application of a model involves forecasting and policy simulation experiments. To generate a forecast with a model the time paths of the exogenous variables have to be predicted, which may be the result of another econometric model forecast or may consist of assumptions about probable policy decisions. The dependence of a model forecast on the exogenous variables constitutes the conditional character of model forecasts. Alternative assumptions about important exogenous variables lead to alternative forecasts and open the way to policy simulation experiments. One can attempt to quantify the effect of policy changes by studying the multiplier effects of exogenous policy variables in a model.

5.0 APPLICATION: A MACRO-ECONOMETRIC MODEL FOR MEXICO

The material presented in the previous sections is illustrated by a macro-econometric model for Mexico. This model is essentially a condensed version of a model built by Abel Beltran del Rio and Lawrence R. Klein[1)] The level of disaggregation is approximately the same in both models, except that this version does not show all the details of the foreign and the fiscal sectors. The model estimated in this section contains many of the several equations outlined in section two.

The model presented here consists of 45 endogenous variables. These are explained by 20 stochastic equations and 25 definitional identities. Exogenous variables include the export sector, import prices, a demographic variable - the rate of population growth - and two important policy instruments, the rates of direct and indirect taxation. Additional exogenous variables are the dollar/peso exchange ratio, and several financial variables, namdy, the amount of domestic and foreign debt of the public sector and foreign reserves. A dummy variable takes care of revisions in the definition of the labor force data.

Stochastic equations explain the following demand categories: private and public consumption, private and public gross fixed investment. Private consumption expenditures are determined by a geometric lag distribution in disposable income. In the public consumption equation this variable is replaced by tax revenues. Adjustments of the capital stock, and the availability of foreign reserves to finance foreign investment goods are the determinants for private gross fixed investment. Tax revenues and the availability

---

1) Abel Beltran del Rio and Lawrence R. Klein: Macroeconometric Model Building in Latin America: The Mexican Case. In: "The Role of the Computer in Economic and Social Research in Latin America", National Bureau of Economic Research, New York 1974.

of domestic and foreign loans to the public sector explain investment level.

Imports of goods are split into consumer goods, capital goods, and raw materials. The main explainatory variable is the corresponding final demand component or domestic economic activity. The effect of import substitution is found in the import equation for capital goods. Imports of services are explained by the capacity to import, the purchasing power of exports. Factor imports are related to domestic economic activity.

Value-added output in the primary, secondary, and tertiary sector is explained by the level of aggregate demand, taking account of the adjustment process of the production sectors with respect to changes in demand. As the export sector is exogenous the identity relating final demand to gross domestic product is solved for inventory change .

Capital formation equations are needed to determine potential production in the rural and urban sector of the economy. These have important feedback both to labor supply (degree of urbanization) and wage rate. The different development of rural and urban labor supply is mainly determined by the different production capabilities caused by a better capital endowment of the urban sector. Whereas further growth of rural production increases the labor participation rate in this sector, the adverse effect is observed in the urban sector.

The average wage rate is explained by the general price level and capacity utilization in the urban sector, expressed as the difference between potential and actual output.

The general price level, the GNP deflator at market prices was corrected for the indirect tax rate. The remaining GNP deflator at factor costs is explained by wage costs per unit of output and import prices, to take into account the openess of the Mexican economy.

Even in a model of moderate size there is some danger of loosing the view of the overall structure of the system. One way to overcome this problem is to draw a block diagram of the relationships among the variables in the system. Another, more concise one, is to represent the model structure by an interaction matrix. The row entries of such an interaction matrix refer to the endogenous variables of a model, the column entries to the following three categories of variables: the simultaneous endogenous, lagged endogenous, and exogenous variables. The interaciton matrix for our Mexican model is contained in tables 3 to 5. It is used as follows: If we are interested, for example, in studying the effects of exchange rate variations, we look at columns 13 to 17 of the exogenous interaction matrix and find direct effects on private and public fixed investment. Indirect effects can be traced in the simultaneous and lagged part of the endogenous interaction matrix. In columns 4 and 5 of the endogenous interaction matrix we recognize effects of private and public investment on total investment expenditures,which in turn, as can be seen in column 7, cause reactions on imports, GNP, and sectoral value output. The corresponding columns of these variables in the endogenous interaction matrix indicate further effects on the endogenous variables. If we continue to follow in this way the reactions of exchange rate variations, finally we see that all endogenous variables will be effected. But this gives only the impact effects. The lagged interaction matrix indicates, that there will be, in addition, dynamic response to a one-time exchange rate variation .

The analysis of a model by studying the interaction matrix may reveal that important relationships between variables are missing, and may again lead to an improvement of the model structure. Mention should be made of the fact that for a linear model the entries of the interaction matrix consist of the model coefficients.

The model was solved by the Gauss-Seidel solution technique. Between 10 and 16 iterations were needed for convergence. Error characteristics of ex-post forecasts between 1952 and 1973 are contained in table 6, giving the mean square error (MSE) and root mean square error (RMSE), and mean absolute percentage error (MAPE) for selected endogenous variables.

	MSE	RMSE	MAPE
Private consumption	6.482	2.546	3.361
Gross national product	15.206	3.900	3.543
GNP deflator	.022	.147	5.245

Table 6: Error Characteristics of Ex-post Forecasts

We complete our example of a macro-econometric model for Mexico with a list of equations and a list of variables.

	1 CP	2 CG	3 C	4 IFP	5 IFG	6 II	7 I	8 X	9 MGC	10 MGI	11 MGR	12 MG	13 MS	14 MF	15 M	16 GNP	17 GNPCUR	18 Q1	19 Q2	20 Q3	21 GDP
CP 1			●						●									●			
CG 2			●																		
C 3						●										●			●	●	
IFP 4						●	●														
IFG 5						●	●														
II 6							●														
I 7										●						●			●	●	
X 8						●							●			●			●	●	
MGC 9												●									
MGI 10												●									
MGR 11												●									
MG 12															●						
MS 13															●						
MF 14						●									●						
M 15						●										●					
GNP 16																	●				
GNPCUR 17																					
Q1 18																					●
Q2 19										●	●			●							●
Q3 20										●	●			●							●
GDP 21				●		●															
KFP 22																					
KFG 23																					
KF1 24																					
KF23 25																					
Q1P 26																					
Q23P 27																					
N 28																					
N23N 29																					
N23 30																					
N1 31																					
LN1 32																					
L1 33																					
LN23 34																					
L23 35																					
L 36																					
TI 37		●			●																
DEP 38																					
Y 39																					
TD 40		●			●																
YD 41	●																				
YWL 42																					
YW 43																					
YNW 44																					
PGNP 45	●	●		●	●								●				●				

Tab. 3.1: Endogenous Interaction Matrix

		22 KFP	23 KFG	24 KF1	25 KF23	26 Q1P	27 Q23P	28 N	29 N23N	30 N23	31 N1	32 LN1	33 L1	34 LN23	35 L23	36 L	37 TI	38 DEP	39 Y	40 TD	41 YD	42 YWL	43 YW	44 YNW	45 PGNP
CP	1																								
CG	2																								
C	3																								
IFP	4	●																							
IFG	5		●																						
II	6																								
I	7																								
X	8																								
MGC	9																								
MGI	10																								
MGR	11																								
MG	12																								
MS	13																								
MF	14																								
M	15																								
GNP	16																								●
GNPCUR	17																●	●	●						
Q1	18											●													
Q2	19													●								●			
Q3	20													●								●			
GDP	21																								●
KFP	22			●	●																				
KFG	23			●	●																				
KF1	24				●																				
KF23	25																								
Q1P	26								●																
Q23P	27								●													●			
N	28									●	●														
N23N	29									●															
N23	30										●				●										
N1	31												●												
LN1	32												●												
L1	33															●									
LN23	34														●										
L23	35															●									
L	36																						●		
TI	37																		●						●
DEP	38																		●						
Y	39																			●	●			●	
TD	40																				●				
YD	41																								
YWL	42																						●		
YW	43																							●	●
YNW	44																								
PGNP	45																					●			

Tab. 3.2: Endogenous Interaction Matrix (continued)

		1 CP	2 CG	3 C	4 IFP	5 IFG	6 II	7 I	8 X	9 MGC	10 MGI	11 MGR	12 MG	13 MS	14 MF	15 M	16 GNP	17 GNPCUR	18 Q1	19 Q2	20 Q3	21 GDP
1	CP(1)	●																				
2	CG(1)		●																			
3	MGC(1)									●												
4	MGI(1)										●											
5	MGR(1)											●										
6	MS(1)													●								
7	MF(1)														●							
8	GNP(1)																					
9	Q1(1)																					
10	Q1(2)																					
11	Q2(1)																			●		
12	Q2(2)																					
13	Q3(1)																				●	
14	Q3(2)																					
15	GDP(1)				●																	
16	GDP(2)				●																	
17	GDP(3)				●																	
18	GDP(4)				●																	
19	KFP(1)				●																	
20	KFG(1)																					
21	KF1(1)																					
22	KF23(1)																					
23	Q1P(1)																					
24	Q1P(2)																					
25	Q23P(1)																					
26	Q23P(2)																					
27	N(1)																					
28	YWL(1)																					
29	YW(1)																					
30	TI(1)					●																
31	TI(2)					●																
32	TI(3)					●																
33	TD(1)					●																
34	TD(2)					●																
35	TD(3)					●																
36	PGNP(1)				●	●																
37	PGNP(2)					●																
38	PGNP(3)					●																
39	PGNP(4)					●																

Tab. 4.1: Lagged Endogenous Interaction Matrix

	22 KFP	23 KFG	24 KF1	25 KF23	26 Q1P	27 Q23P	28 N	29 N23N	30 N23	31 N1	32 LN1	33 L1	34 LN23	35 L23	36 L	37 TI	38 DEP	39 Y	40 TD	41 YD	42 YWL	43 YW	44 YNW	45 PGNP	
CP(1)																									1
CG(1)																									2
MGC(1)																									3
MGI(1)																									4
MGR(1)																									5
MS(1)																									6
MF(1)																									7
GNP(1)											●													●	8
Q1(1)											●														9
Q1(2)																									10
Q2(1)													●												11
Q2(2)													●												12
Q3(1)													●												13
Q3(2)													●												14
GDP(1)																								●	15
GDP(2)																									16
GDP(3)																									17
GDP(4)																									18
KFP(1)	●																								19
KFG(1)		●																							20
KF1(1)					●																				21
KF23(1)						●																			22
Q1P(1)								●																	23
Q1P(2)								●																	24
Q23P(1)								●																	25
Q23P(2)								●																	26
N(1)							●																		27
YWL(1)																					●				28
YW(1)																								●	29
TI(1)																								●	30
TI(2)																									31
TI(3)																									32
TD(1)																									33
TD(2)																									34
TD(3)																									35
PGNP(1)																					●			●	36
PGNP(2)																									37
PGNP(3)																									38
PGNP(4)																									39

Tab. 4.2: Lagged Endogenous Interaction Matrix (continued)

	1 CP	2 CG	3 C	4 IFP	5 IFG	6 II	7 I	8 X	9 MGC	10 MGI	11 MGR	12 MG	13 MS	14 MF	15 M	16 GNP	17 GNPCUR	18 Q1	19 Q2	20 Q3	21 GDP
XG 1																					
XS 2								●													
XF 3								●													
SBS 4						●		●								●					
SBF 5																●					
PM 6													●								
PM(1) 7																					
NG 8																					
DUML 9																					
RTI 10																					
RTD 11																					
RDEP 12																					
REX 13				●	●																
REX(1) 14				●	●																
REX(2) 15					●																
REX(3) 16					●																
REX(4) 17					●																
LGD 18					●																
LGD(1) 19					●																
LGD(2) 20					●																
LGD(3) 21					●																
LGD(4) 22					●																
LGF$ 23					●																
LGF$(1) 24					●																
LGF$(2) 25					●																
LGF$(3) 26					●																
LGF$(4) 27					●																
RF$ 28				●																	
RF$(1) 29				●																	
TIME 30																					

Tab. 5.1: Exogenous Interaction Matrix

		22 KFP	23 KFG	24 KF1	25 KF23	26 Q1P	27 Q23P	28 N	29 N23N	30 N23	31 N1	32 LN1	33 L1	34 LN23	35 L23	36 L	37 TI	38 DEP	39 Y	40 TD	41 YD	42 YWL	43 YW	44 YNW	45 PGNP
XG	1																								
XS	2																								
XF	3																								
SBS	4																								
SBF	5																								
PM	6																								●
PM(1)	7																								●
NG	8							●																	
DUML	9											●		●											
RTI	10																●								
RTD	11																			●					
RDEP	12																	●							
REX	13																								
REX(1)	14																								
REX(2)	15																								
REX(3)	16																								
REX(4)	17																								
LGD	18																								
LGD(1)	19																								
LGD(2)	20																								
LGD(3)	21																								
LGD(4)	22																								
LGF$	23																								
LGF$(1)	24																								
LGF$(2)	25																								
LGF$(3)	26																								
LGF$(4)	27																								
RF$	28																								
RF$(1)	29																								
TIME				●																					

Tab. 5.2: Exogenous Interaction Matrix (continued)

LIST OF EQUATIONS

A. AGGREGATE DEMAND

A.1 Private Consumption

$$CP = \underset{(.11231)}{.20067}\ CP(1) + \underset{(.08511)}{.64984}\ YD/PGNP + \underset{(.81393)}{4.65384}$$

R^2=.999 S.E.=1.01012 DW=1.88 1950-73

A.2 Public Consumption

$$CG = \underset{(.14765)}{.81658}\ CG(1) + \underset{(.11887)}{.23919}\ (TI+TD)/PGNP - \underset{(.28382)}{.53434}$$

R^2=.992 S.E.=.35494 DW=1.65 1950-73

A.3 Consumption

$$C = CP + CG$$

A.4 Private Gross, Fixed Investment

$$\begin{aligned} IFP = {} & \underset{(.01297)}{.06830}\ KFP(1) + \underset{(.05622)}{.24494}\ (GDP-GDP(1)) + \\ & + \underset{(.04718)}{.41938}\ (GDP(1)-GDP(2)) + \underset{(.05063)}{.43245}\ (GDP(2)-GDP(3)) + \\ & + \underset{(.07580)}{.28415}\ (GDP(3)-GDP(4)) + \underset{(.23614)}{.80759}\ (RF\$ \times REX/PGNP- \\ & -RF\$(1) \times REX(1)/PGNP(1)) - \underset{(.25536)}{1.01284} \end{aligned}$$

R^2=.992 S.E.=.46147 DW=2.37 1952-73

A.5 Public Gross Fixed Investment

$$IFG = \underset{(.05137)}{.13221}\ FG + \underset{(.00973)}{.19830}\ FG(1) + \underset{(.03542)}{.19830}\ FG(2) + \underset{(.03278)}{.13220}\ FG(3) + \underset{(.26049)}{.23190}$$

R^2=.975 S.E.=.54804 DW=1.37 1952-73

$$FG = TI/PGNP + TD/PGNP + LGD/PGNP - LGD(1)/PGNP(1) + LGF\$/PGNP - LGF\$(2)/PGNP(2)$$

A.6 Inventory Changes

$$II = GDP - C - IFP - IFG - X + M + XF - MF + SBS$$

A.7 Investment

$$I = IFP + IFG + II$$

A.8 Exports of Goods, Services and Factors

$$X = XG + XS + XF$$

A.9 Imports of Consumer Goods

$$MGC = .64246\ MGC(1) + .00785\ CP + .04100$$
$$\quad (.22186) \qquad (.00349) \qquad (.14006)$$

R^2=.845 S.E.=.19625 DW=1.54 1950-73

A.10 Imports of Capital Goods

$$MGI = .22460\ MGI(1) + .14707\ I + -.02211\ (Q2+Q3) + 1.89633$$
$$\quad (.19002) \qquad (.07712) \qquad (.01853) \qquad (.50133)$$

R^2=.859 S.E.=.30511 DW=1.81 1950-73

A.11 Imports of Raw Materials & Fuels

$$MGR = .81383\ MGR(1) + .00531\ (Q2+Q3) + .20893$$
$$\quad (.25781) \qquad (.00272) \qquad (.54559)$$

R^2=.693 S.E.=.31604 DW=1.21 1950-73

A.12 Imports of Goods

$$MG = MGC + MGI + MGR$$

A.13 Imports of Services

$$MS = .55963\ MS(1) + .14145\ (X \times PGNP)/PM - .46773$$
$$\quad (.18538) \qquad (.05862) \qquad (.28411)$$

R^2=.960 S.E.=.22416 DW=1.78 1950-73

A.14 Imports of Production Factors

$$MF = \underset{(.17251)}{.43638}\ MF(1) + \underset{(.00323)}{.01232}\ (Q2+Q3) - \underset{(.05233)}{.11161}$$

$R^2 = .983$ S.E. = .09806 DW = 1.83 1950-73

A.15 Imports of Goods, Services and Factors

$$M = MG + MS + MF$$

A.16 Gross National Product

$$GNP = C + I + X - M - SBS - SBF$$

A.17 Gross National Product

$$GNPCUR = GNP \times PGNP$$

B. VALUE - ADDED OUTPUT

B.1 Primary Production

$$Q1 = \underset{(.01203)}{.08214}\ CP + \underset{(.86975)}{8.40382}$$

R^2=.665 S.E.=1.52457 DW=.32 1950-73

B.2 Secondary Production

$$Q2 = \underset{(.16556)}{.54660}\ Q2(1) + \underset{(.04882)}{.15285}\ (C+I+X) - \underset{(.70954)}{1.01215}$$

R^2=.992 S.E.=1.17462 DW=1.98 1950-73

B.3 Tertiary Production

$$Q3 = \underset{(.14152)}{.60193}\ Q3(1) + \underset{(.07418)}{.25214}\ (C+I+X) - \underset{(1.63347)}{4.91467}$$

R^2=.995 S.E.=1.83381 DW=1.58 1950-73

B.4 Gross Domestic Product

$$GDP = Q1 + Q2 + Q3$$

C. CAPITAL FORMATION

C.1 Private Capital Stock

KFP = IFP + .90 KFP(1)

C.2 Public Capital Stock

KFG = IFG + .95 KFG(1)

C.3 Capital Stock in Rural Sector

KF1 = .05946 (KFP+KFG) - .18022 TIME + 12.21463
(.00345) (.02822) (1.38097)

R^2=.988 S.E.=.24944 DW=1.53 1950-73

C.4 Capital Stock in Urban Sector

KF23 = KFP + KFG - KF1

D. POTENTIAL PRODUCTION

D.1 Rural Capacity

$$Q1P = 1.84673\ KF1(1) + 2.35746$$
$$(.15758) \qquad (1.17716)$$

R^2=.856 S.E.=1.50689 DW=.15 1950-73

D.2 Urban Capacity

$$Q23P = .72630\ KF23(1) + 12.33340$$
$$(.01331) \qquad (1.40187)$$

R^2=.993 S.E.=3.27641 DW=.38 1950-73

E. LABOR SUPPLY

E.1 Population

$$N = NG \times N(1)$$

E.2 Urbanization: N23/N

$$N23N = \underset{(.0004)}{.00052}\,(Q23P-Q1P) + \underset{(.00005)}{.00069}\,(Q23P(1)-Q1P(1)) + \underset{(.00005)}{.00052}\,(Q23P(2)-Q1P(2)) + \underset{(.00881)}{.41702}$$

$R^2=.885$ S.E.=.01980 DW=.10 1950-73

E.3 Urban Population

$$N23 = N23N \times N$$

E.4 Rural Population

$$N1 = N - N23$$

E.5 Rural Labor Participation Rate

$$LN1 = \underset{(.00014)}{.00209}\ Q1 + \underset{(.00019)}{.00279}\ Q1(1) + \underset{(.00014)}{.00209}\ Q1(2) - \underset{(.00259)}{.01437}\ DUML + \underset{(.00597)}{.26112}$$

$R^2=.909$ S.E.=.00520 DW=.80 1950-73

E.6 Rural Labor Force

$$L1 = LN1 \times N1$$

E.7 Urban Labor Participation Rate

$$LN23 = -\underset{(.00002)}{.00013}\ (Q2+Q3) - \underset{(.00003)}{.00017}\ (Q2(1)+Q3(1)) - \underset{(.00002)}{.00013}\ (Q2(2)+Q3(2)) + \underset{(.00444)}{.01357}\ DUML + \underset{(.00514)}{.30030}$$

$R^2=.728$ S.E.=.01049 DW=.61 1950-73

E.8 Urban Labor Force

$$L23 = LN23 \times N23$$

E.9 Labor Force

$$L = L1 + L23$$

F. INCOME AND PRICES

F.1 Indirect Taxes

$$TI = RTI \times GNPCUR$$

F.2 Depreciation

$$DEP = RDEP \times GNPCUR$$

F.3 National Income

$$Y = GNPCUR - TI - DEP$$

F.4 Direct Taxes

$$TD = RTD \times Y$$

F.5 Disposable Income

$$YD = Y - TD$$

F.6 Wage Rate

$$YWL = (\underset{(.13067)}{1.65179}\,(PGNP-PGNP(1))/PGNP(1)\times 100. - \underset{(.21433)}{.22012}\,(Q23P-Q2-Q3) + \underset{(1.49527)}{1.40860})\times YWL(1)/100. + YWL(1)$$

$R^2=.916$ S.E.=2.28332 DW=2.01 50-73

F.7 Wage Income

$$YW = YWL\times L$$

F.8 Non-Wage Income

$$YNW = Y - YW$$

F.9 General Price Level:GNP Deflator

$$PGNP = (\underset{(.07936)}{.46415}\,(YW/GDP - YW(1)/GDP(1))/(YW(1)/GDP(1))\times 100. + \underset{(.08693)}{.23199}\,(PM-PM(1))/PM(1)\times 100. + \underset{(.29759)}{1.65207})\times(PGNP(1) - TI(1)/GNP(1))/100. + TI/GNP + PGNP(1) - TI(1)/GNP(1)$$

$R^2=.945$ S.E.=1.00145 DW=2.19 50-73

LIST OF VARIABLES

C	Consumption	Bill.1950 Pesos	
CG	Public Consumption	Bill.1950 Pesos	
CP	Private Consumption	Bill.1950 Pesos	
DEP	Depreciation	Bill.Curr.Pesos	
DUML	Dummy: Revisions in Labor Data	1960-68 = 1	EXOG.
GDP	Gross Domestic Product	Bill.1950 Pesos	
GNP	Gross National Product	Bill.1950 Pesos	
GNPCUR	Gross National Product	Bill.Curr.Pesos	
I	Investment	Bill.1950 Pesos	
IFG	Public Gross Fixed Investment	Bill.1950 Pesos	
IFP	Private Gross Fixed Investment	Bill.1950 Pesos	
II	Inventory Changes	Bill.1950 Pesos	
KFG	Public Capital Stock	Bill.1950 Pesos	
KFP	Private Capital Stock	Bill.1950 Pesos	
KF1	Capital Stock in Rural Sector	Bill.1950 Pesos	
KF23	Capital Stock in Urban Sector	Bill.1950 Pesos	
L	Labor Force	Mill.	
LGD	Public Domestic Debt	Bill.Curr.Pesos	EXOG.
LGF$	Public Foreign Debt	Mill.Curr.Dollar	EXOG.
LN1	Rural Labor Participation Rate	Mill.	
LN23	Urban Labor Participation Rate	Mill.	
L1	Rural Labor Force	Mill.	
L23	Urban Labor Force	Mill.	
M	Imports of Goods,Services,Factors	Bill.1950 Pesos	
MF	Imports of Productive Factors	Bill.1950 Pesos	
MG	Imports of Goods	Bill.1950 Pesos	

MGC	Imports of Consumption Goods	Bill.1950 Pesos	
MGI	Imports of Investment Goods	Bill.1950 Pesos	
MGR	Imports of Raw Materials	Bill.1950 Pesos	
MS	Imports of Services	Bill.1950 Pesos	
N	Population	Mill.	
NG	Rate of Population Growth	N/N(1)	EXOG.
N1	Rural Population	Mill.	
N23	Urban Population	Mill.	
N23N	Urbanization	Mill.	
PGNP	General Price Level	1950 = 100	
PM	Import Price Index	1950 = 100	EXOG.
Q1	Primary Production	Bill.1950 Pesos	
Q1P	Rural Capacity	Bill.1950 Pesos	
Q2	Secondary Production	Bill.1950 Pesos	
Q23P	Urban Capacity	Bill.1950 Pesos	
Q3	Tertiary Production	Bill.Curr.Pesos	
RDEP	Rate of Depreciation	DEP/GNPCUR	EXOG.
REX	Rate of Exchange	$/PESOS	EXOG.
RF$	Foreign Reserves	Mill.Curr.Dollar	EXOG.
RTD	Rate of Direct Taxes	TD/Y	EXOG.
RTI	Rate of Indirect Taxes	TI/GNPCUR	EXOG.
SBF	Stat.Discr.Balance of Factors	Bill.1950 Pesos	EXOG.
SBS	Stat.Discr.Balance of Services	Bill.1950 Pesos	EXOG.
TD	Direct Taxes	Bill.Curr.Pesos	
TI	Indirect Taxes	Bill.Curr.Pesos	
TIME	Time	1948 = 48	EXOG.
X	Exports of Goods, Services, Factors	Bill.1950 Pesos	
XF	Exports of Productive Factors	Bill.1950 Pesos	EXOG.
XG	Exports of Goods	Bill.1950 Pesos	EXOG.
XS	Exports of Services	Bill.1950 Pesos	EXOG.

Y	National Income	Bill.Curr.Pesos
YD	Disposable Income	Bill.Curr.Pesos
YNW	Non-Wage Income	Bill.Curr.Pesos
YW	Wage Income	Bill.Curr.Pesos
YWL	Wage Rate	Thous.Curr.Pesos