Arun Shourie

The Use of Macro-economic Regression Models of Developing Countries for Forecasts and Policy Prescription

Reprinted from *Oxford Economic Papers* Volume 24, No. 1, March 1972
THE USE OF MACRO-ECONOMIC REGRESSION MODELS OF DEVELOPING COUNTRIES FOR FORECASTS AND POLICY PRESCRIPTION: SOME REFLECTIONS ON CURRENT PRACTICE

By ARUN SHOURIE

I. Introduction

The need for long-term forecasting of various magnitudes in underdeveloped economies is increasingly recognized. Lending agencies that make loans with maturities ranging from ten to fifty years naturally want to know all they can about likely developments in these economies for as many years into the future as possible. An international agency like UNCTAD which has learned that it takes a good many years to wrest action from a hundred-odd countries must plan now for steps it would want taken in the mid or late 1970s, and to do so it needs projections relating to those distant periods. National planning agencies often want to examine the problems of their economies in a long-term context and, just as often, they must put together plan documents that give a long-term perspective to current developments, their programmes, and their requests for external assistance.

It is a common practice in making these forecasts, whether medium- or long-term, to use econometric models that were originally developed for short-term forecasting of relationships in developed economies. Thus UNCTAD has a large project under way for constructing some thirty models of various economies; ECAFE recently published econometric models for ten countries; other international organizations, such as the IMF and the World Bank, and national donor agencies, such as the U.S. AID, are making increasing use of them; and consultants and research organizations working on problems of the third world often base their projections and advice on econometric models.

This paper is based on a review of about forty-five such models that have been proposed for various developing countries. All of these models use statistically estimated parameters to provide a quantitative description of certain macro-relationships in developing economies. They might be described as aggregative macro-economic regression models.

Sections II to VI review the models from five points of view: the reliability

1 I am grateful to P. D. Henderson, B. S. Minhas, E. P. Holland, N. G. Carter, and R. Cheatham for many useful comments and to S. Malik and E. Ramiscal for their help in the computations. None of them is responsible for any errors that remain. The author is a member of the Economics Department of the World Bank. The views expressed in this paper are his personal views and not those of the World Bank group.
of the data from which they are estimated; the significance of the coefficients of their equations; the functional form of their equations; their propensity to use certain statistical tests in a mechanical way; and their use for making forecasts and prescribing policies for years far into the future. Section VII summarizes the principal conclusions and contains some general remarks about the manner in which econometric models are being used at present.

The discussion in Sections II to VI is illustrated by cases from various models that have been proposed in the last five years and from a conventional two-gap model of the Kenya economy put together by the author. Many of the illustrations are taken from a recent UNCTAD study that contains models for eighteen countries. Raul Prebisch in his preface states that ‘the study provides a solid basis for the elaboration of new national and international targets for the United Nations Development Decade of the 1970’s’.

II. The data base

It may appear trite or even superfluous to begin yet another article on econometric models by drawing attention to the data from which the models are estimated. But the fact is that reliable and ample data is critical to an econometric model: if the underlying data is seriously deficient then our care in adopting optimal methods of estimation is of little avail. Every economist constructing a model is conscious of this fact and acknowledges the problems connected with the basic data by some ritualistic disclaimers: the data are of uncertain quality but this is the best that is available; the alternative to using it is to abandon all hopes of quantifying policy recommendations; in any case the exercise is worth while for it highlights the particular areas in which the data-collection efforts need to be strengthened. Having protected his flanks with these disarming

1 UNCTAD, Trade Prospects and Capital Needs of Developing Countries, United Nations, New York, 1968 (henceforth, UNCTAD, 1968). A similar volume has been published by the Economic Commission for Asia and the Far East: ECAFE, Feasible Growth and Trade Gap Projections in the ECAFE Region, Bangkok, 1968 (henceforth, ECAFE, 1968). This volume contains models for nine countries. The model for Thailand, however, is the same as the one contained in UNCTAD, 1968. All subsequent references to ECAFE, 1968 relate to the eight models other than the model for Thailand.


3 J. T. S. and Earl O. Heady, ‘Econometricians and the data gap; reply’, American Journal of Agricultural Economics, vol. 51, No. 1, Feb. 1969, p. 188. UNCTAD, 1968 notes ‘the paucity, inadequacy and unreliability of the relevant statistical data (for, in this case, Nigeria). . . . Suffice it to say that in many cases the estimates are rather crude and impressionistic with very little firm basis . . . ’, p. 180. It goes on to note that the national accounts in Nigeria are intimately tied up with population estimates as for many items the aggregate magnitude is obtained by multiplying an estimated per capita figure by the population estimate and that the 1963 census has shown population to be much higher than the figures assumed in the national accounts. ‘Nevertheless, use of available data is probably better than not using any data at all and provides a useful check on more qualitative approaches’, ibid.
assertions the model-builder often proceeds to act in precisely the way he would have acted had the data been abundant and reliable.

The data of most developing countries present three problems to the econometrician.

First, the series for most macro-variables cover only a few years. Consequently, from an econometric point of view the models operate on the subsistence margin in that their relationships are estimated from no more than five to sixteen observations.¹

Second, the figures are of varying and often very low reliability. Estimates of many macro-variables are based on guesses or on rules of thumb derived from scanty empirical evidence. Investigators who have had the patience to examine the data carefully have often concluded that 'we may have to admit that to now we really are not in a position to tell how the main economic variables are changing in large areas of the third world'.² For these countries trade statistics, in fact, constitute the only data in which most economists are willing to place a good deal of confidence. But even this information when it is examined systematically—for example, when it is compared with the returns of the country’s trading partners—reveals very substantial discrepancies.³

The underlying data are such that many refinements in estimation procedures, such as multiple stage least squares, seem to be somewhat pointless. In fact, one does not quite know how to assess even the more basic refinements. Consider the fact that most models in use base their estimates on constant price data.⁴ Relationships are estimated at constant prices so as to separate the income and price effects. More often than not the model builder adopts the national accounts data for a country without examining the deflators used to arrive at the constant price series. To cite one instance: UNCTAD, 1968 gives projections for Ceylon at constant 1960 prices.⁵ Production in the agricultural sector consists of a few

¹ The frequency distribution of the models in UNCTAD, 1968 and ECAFE, 1968 by the number of observations on which they are based is as follows:

<table>
<thead>
<tr>
<th>Number of observations</th>
<th>5–7</th>
<th>8–10</th>
<th>11–13</th>
<th>14–16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of countries:</td>
<td>(i)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>UNCTAD, 1968</td>
<td>2</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>ECAFE, 1968</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Equations in the four models of UNCTAD, 1968 that have not been included above are based on varying numbers of observations: Argentina (12 to 16), Ghana (10 to 16), Nigeria (10 to 13), and Philippines (9 to 14).


³ As a recent and relevant illustration, cf. S. Naya and T. Morgan, The Accuracy of International Trade Data: The Case of Southeast Asian Countries, University of Wisconsin, mimeographed, Aug. 1967.

⁴ Fourteen of the eighteen models in UNCTAD, 1968 use constant price series.

⁵ UNCTAD, 1968, p. 75.
identifiable crops: as the product is homogeneous over space and time output indices are used and we do, in fact, obtain a constant price series that is meaningful. For the other sectors value added at current prices is adjusted by the following odd assortment:

<table>
<thead>
<tr>
<th>Sector</th>
<th>Deflator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mining and quarrying</td>
<td>Wage index for workers in industry and commerce.</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>Indis of physical output.</td>
</tr>
<tr>
<td>(a) Tea, rubber, coconut processing</td>
<td>Domestic group of the Colombo cost of living index.</td>
</tr>
<tr>
<td>(b) Factory and cottage industries</td>
<td>Domestic group of the Colombo cost of living index.</td>
</tr>
<tr>
<td>excluding (a)</td>
<td></td>
</tr>
<tr>
<td>Construction</td>
<td>Wage index of Government technical and clerical employees.</td>
</tr>
<tr>
<td>Electricity, water, and sanitary</td>
<td>A mixture of indices of wages of Central Government employees, workers</td>
</tr>
<tr>
<td>services</td>
<td>in industry and commerce, and the Colombo cost of living index.</td>
</tr>
<tr>
<td>Transport, storage, and communication</td>
<td></td>
</tr>
<tr>
<td>Wholesale and retail trade</td>
<td>Domestic group of the Colombo cost of living index.</td>
</tr>
<tr>
<td>Banking, insurance, and real estate</td>
<td>Wage index for employees in industry and commerce.</td>
</tr>
<tr>
<td>Ownership of dwellings</td>
<td>Domestic group of the Colombo cost of living index.</td>
</tr>
<tr>
<td>Services</td>
<td>Colombo cost of living index.</td>
</tr>
</tbody>
</table>

Weights for the more widely used indices are based on surveys conducted many years ago when the structure of the economy was very different from what it is now. The most widely used deflator—the Colombo cost of living index—for instance, is based on a survey conducted in 1952.¹ The problems are compounded for economies that in addition to having deflators of somewhat limited reliability have had multiple or oft-changing exchange rates.²

For a similar reason one must be wary of reading too much into high coefficients of determination and t-ratios before examining the manner in which the relevant data were estimated in the first place. If, for instance, we fit an equation relating imports of construction materials to value added

¹ Some studies explicitly acknowledge that for their countries current price series are more reliable than available constant price series: cf. for Thailand, UNCTAD, 1968, pp. 388, 390; for Tanzania, ibid., p. 190; for Kenya, ibid., p. 107; for Uganda, ibid., pp. 233-4.
by construction in Ceylon we obtain a very satisfactory fit and a highly significant t-ratio. The reason for this is simply that value added by the bulk of the construction sector is itself estimated by multiplying the imports of construction materials by (4*25)—after allowing for a three-month lag. The \( R^2 \) and the t-ratios in this case would tell us no more than the national accounts themselves.\(^1\)

The third source of difficulties is that the national accounts of many of the countries for which econometric models are being constructed at present are in their infancy and are still being revised periodically. Therefore, each of the models is 'dated' in the sense of being tied to the national income series that were in use at the time when it was estimated. A rough check indicates that of the twenty-two countries for which UNCTAD, 1968 provides detailed models the accounts for as many as eleven have been revised since 1965 (the last year for the sample period of UNCTAD, 1968) in ways that would materially affect the estimates of some important coefficients. The national accounts of Kenya, for instance, have been revised often and very considerably over the last few years. One has only to compare the data used by Paul Clark in his models for the East African economies,\(^2\) that used by Faaland and Dahl in their model of Kenya,\(^3\) and the data in use now, to see that series like gross domestic product at factor cost and gross investment have been revised upward by 15 to 20 per cent.

Clearly, national accounts have to be revised as and when new data or information become available. However, these revisions necessarily reduce confidence in models estimated at an earlier stage of the country's experience of preparing national accounts. Consider the consequences of revising capital formation figures upwards by 15 or 20 per cent: the savings figures in Kenya, as in many other countries, are derived as a residual using the capital formation and balance of payments accounts. Now, the capital formation figures have been revised upwards by one-eighth to one-fifth but the balance of payments account has not undergone a revision of any comparable magnitude. Thus today one gets very different savings coefficients and a very different picture of the accelerations or decelerations in the country's savings efforts in a given period than one would get using

---

\(^1\) These remarks are suggested by some equations in a recently published model of the Colombian economy: K. Marwah, 'An econometric model of Colombia: a prototype devaluation view', *Econometrica*, Apr. 1969, pp. 228–51. We are told that 'there seems to be a strong complementary relation between the imports of capital goods and investment. A marginal increment of one billion pesos worth of imports of capital goods seems to be accompanied by an addition of 1.49 billion pesos of total investment', ibid., p. 238. The t-ratio is more than 16 and the \( R^2 \) is 0.981. In Colombia fixed investment in sectors other than construction is estimated by applying multipliers to imported capital goods assigned to the sectors. These multipliers have been changed only once in the last twelve years.


the old series on capital formation. The same holds true of functions linking imports of goods to GDP at factor cost or to fixed investment. And there is no assurance that no further revisions will be needed.

These revisions affect not only individual coefficients but also the overall conclusions that emerge from the models. The Faaland–Dahl model for Kenya, for example, projects the savings gap to be the dominant one and proposes that Kenya should aim at increasing its marginal savings ratio to 25 per cent in the 1970s. The revised accounts being used now indicate, however, that the marginal savings rate in the period 1960–8 itself was 25.2 per cent in contrast to the 11.3 per cent that emerged from the accounts Faaland and Dahl used. Mechanical projections from the author’s Kenya model—a model that in its specification is quite similar to the Faaland–Dahl model—indicate that savings are not likely to be an important issue in the 1970s and that the trade gap rather than the savings gap is liable to be the dominant one.

The equations contained in the kind of models we are now considering do provide a compact description of large amounts of data. As such they may sometimes highlight its peculiarities and may dramatize the fact that the assumptions on which the data has been put together need to be re-examined. Sometimes such insights are obtained as by-products of a model-building exercise; but clearly they cannot by themselves be a sufficient justification for constructing elaborate models out of shaky data. Of course, in cases where the underlying relationships are ‘mechanical’ or ‘biological’ and where they have been extensively documented for a number of similar systems, models can be used to generate data from only a few apparently unrelated bits of information. Demographic models of human and animal populations exemplify the use of models for this purpose. But given our limited understanding of change in developing economies and the even more limited extent to which their experience has been quantified, one can scarcely hope to use these simple and highly aggregated models to generate data just yet.

As we know little about the nature of the errors it is difficult to use methods appropriate to an errors-in-variables situation. Moreover, the results of recent revisions of national accounts dissuade us from assuming too readily that the errors in each variable have a zero mean, that they are independent of the variable, of other variables that figure in the equation, and of the errors in other variables. Given the way in which the national accounts of many countries are put together these requirements are seldom met. Estimates for the series of an individual aggregate variable are derived from diverse sources of which some sketchy information about other variables is certainly an important one. Not only is one

estimate derived or 'built up' from the others but in the final stages they are all adjusted and doctored for securing consistency.

III. The significance of individual coefficients

Even if the data was entirely reliable and covered many years, the coefficients of the regression equations of these models would have only a very limited significance. That this is often overlooked is shown by the fact that in addition to their use for forecasting these highly aggregated models are almost invariably used for purposes of control—i.e. for making policy recommendations. The coefficients of a fitted regression equation are treated not merely as coefficients that have emerged from a particular system of simultaneous equations, which is in fact all that they are, but are too readily identified with the marginal propensities and elasticities to which the economist is accustomed and which have traditionally had overtones of causality around them.

The abuse of regression analysis in this particular form is almost universal and a very large number of examples can be listed. A few will have to suffice.

Thus, for example, a model for Argentina fits private consumption expenditures to wage and non-wage income and gets the equation:

Private consumption expenditure

\[ \text{Private consumption expenditure} = -106.04 + 1.006 \times \text{wage income} + 0.655 \times \text{non-wage income} \]

with an $R^2$ of 0.93. Figures in parentheses at $t$-ratios. It goes on to assert that 'as may be seen from the above equations, there exists a markedly different consumption behaviour between wage and non-wage income groups. Thus, changes in the distribution of income are bound to have a significant effect on the level of total consumption.' The coefficients seem 'reasonable' only because we are accustomed to thinking that the marginal propensity to consume out of 'wage incomes' is much higher than out of 'non-wage incomes' primarily because we associate the former with the poorer and the latter with the richer sections of the population. In fact, in this model most of the income originating in the agricultural sector is treated as 'non-wage income': a fact that must bring the per capita incomes of the wage and non-wage groups closer together. But far more important from our point of view is the fact that the two independent variables are highly collinear.\(^2\) It is only fortuitous that the coefficient for 'non-wage income' has turned out to be smaller than that for 'wage income'.

\(^1\) UNCTAD, 1968, p. 408.

\(^2\) The simple correlation between wage and non-wage incomes at current prices is 0.9882 (cf. data in UNCTAD, 1968, p. 427). The equation is based on incomes at constant prices. It has not been possible to obtain the correlation coefficient for these as the report does not furnish the data on incomes at constant prices.
Similarly, a recently published model of the Colombian economy designed specifically to analyse the consequences of changes in exchange-rate policies treats each of the coefficients as economic propensities and elasticities—even when some important explanatory variables that are left out are highly collinear with the ones that have been included.¹

B. A. de Vries and J. C. Liu in their recent paper on Brazil² provide a very explicit illustration of economists attributing causal significance to each of their coefficients. At each step this instructive paper carefully stresses the number of significant explanatory variables that have not been included in each equation (many of these are probably collinear with the ones that have been included), the poor quality of data, the partial specification of the model itself, and many other difficulties; and yet it interprets the resulting coefficients as if they were the conventional elasticities and propensities to which the economist is accustomed. The paper proceeds to derive trade-offs between inflation and growth.

For a large number of reasons we are not justified in using regression equations and their coefficients for control purposes. We shall discuss two of these reasons. First, there is the fact that in a very large number of cases the explanatory variables that have been included in an equation are highly collinear and, second, the likelihood that the equations are misspecified and that some of the excluded variables may well be collinear with the variables that have been included in the equation.

(a) Collinearity among regressors³

We know that if the regressors are orthogonal, if the equation is specified correctly, and if the data is reliable, an unambiguous meaning can be attached to the estimated coefficients. At the other extreme, the coefficients cannot be estimated if the regressors are perfectly collinear. In practice the situation encountered is an intermediate one: over short periods of five to fifteen years a surprisingly large number of aggregative time series have strong trend components and are, therefore, highly though not perfectly collinear.⁴ The use of such variables as independent regressors in multiple regression equations raises difficult problems of interpretation.

¹ K. Marwah, op. cit., in particular, the discussion of the consumption equations, pp. 234-5.
² B. A. de Vries and J. C. Liu, op. cit.
⁴ Of the 630-odd estimated equations that are listed in UNCTAD, 1968 about 168 have more than one regressor. The data given in UNCTAD, 1968 enabled us to calculate the
Individual coefficients have little significance if the regressors in a multiple regression equation are collinear. In particular if the regressors $X_1$ and $X_2$ are not causally independent it is misleading to assume that a coefficient, say $b_2$, in the equation $(Y = b_1 X_1 + b_2 X_2 + u)$ indicates the extent to which $Y$ will change as a consequence of a unit change in $X_2$ 'other things remaining the same'. To begin with, the high correlation between $X_1$ and $X_2$ over the sample period may imply that there is some organic dependence between the two regressors so that when $X_2$ is altered 'other things'—in this case, $X_1$—may not 'remain the same'. The coefficient $b_2$ in the single equation model $(Y = b_1 X_1 + b_2 X_2 + u)$ does not take account of these indirect effects of $X_2$ on $Y$ via the changes it induces in $X_1$. Even if there is reason to believe that there is no organic dependence between $X_1$ and $X_2$ and that they 'just happened to have been collinear in the sample period' there are various reasons to be wary of using the coefficients as if they were the partial derivatives of calculus.

First, if the correlations between regressors are indeed as high as (0.95) (as they most often are in the models under review—(0.9985) in Tanzanian equation cited below) the values obtained for the coefficients are very sensitive to sample coverage, data errors, specification of the equation, and even rounding-off errors. The argument can be stated compactly as follows. Let $R$ be the matrix of coefficients of correlation ($r_{ij}$) between the regressor variables; let $b^*$ denote the vector of standard partial regression coefficients so that the regression coefficient $b_j$ is equal to $(b_j^* S_j/S_{X_j})$ and let $g$ be the vector of coefficients of correlation between the dependent variable and the regressors. Then,

$$\begin{align*} R b^* &= g, \\ b^* &= R^{-1} g, \end{align*}$$

Now, $R^{-1}$ is $[(r_{ij})'/|R|]$ where $[R_{ij}]$ is the adjoint matrix of $R$ consisting of cofactors and $|R|$ is the determinant obtained from $R$. The closer a regressor is to being the linear function of other regressors the closer the value of the determinant $|R|$ will be to zero. With a divisor close to zero, the values of elements in the inverse matrix $R^{-1}$ (from which the vector $b^*$ and ultimately $b$ and the standard error of the individual coefficients are estimated) become extremely volatile: even small changes in the $r_{ij}$ arising intercorrelations among the regressors in 118 of these equations. A total of 168 intercorrelations among different regressors were involved. The distribution of these intercorrelations is as follows:

<table>
<thead>
<tr>
<th>Value of the correlation coefficient</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0 to 0.6</td>
<td>12</td>
</tr>
<tr>
<td>0.65 &gt; 0.6</td>
<td>6</td>
</tr>
<tr>
<td>0.7 &gt; 0.75</td>
<td>6</td>
</tr>
<tr>
<td>0.8 &gt; 0.85</td>
<td>12</td>
</tr>
<tr>
<td>0.9 &gt; 0.95</td>
<td>4</td>
</tr>
<tr>
<td>1.0 &gt; 1.0</td>
<td>13</td>
</tr>
</tbody>
</table>

In other words, about 60 per cent of the intercorrelations among supposedly 'independent' variables are higher than 0.95 and about 70 per cent are higher than 0.90.
from sample coverage, or data errors or from the effects of rounding-off
can alter the value of \(|R|\) greatly. In order to gauge the extent of this
volatility it is sufficient to examine the simple correlation between the
regressors when the equation has only two regressors. But if there are
more than two regressors the value of \(R\) can be close to zero even if none
of the correlation coefficients is individually close to 1.0. The reason for
this is that an individual regressor \(X_i\) though not very highly correlated
with any one of the other regressors may yet be a linear function of all or
some (i.e. more than one) of the other regressors. In this case an inspection
of the simple correlation coefficients will not suffice; one must look at the
value of the determinant \(|R|\) itself.\(^1\)

Second, the standard error of the estimate of a coefficient increases as
the degree of collinearity increases. When the correlations among regressors
are as high as 0.9985 or 0.95 the standard errors are often so high as to make
the point estimate of the coefficient virtually useless for policy purposes.
In this context two conventions regarding standard errors and \(t\)-ratios
deserve comment. First, the customary rule of thumb that is used in
evaluating the \(t\)-ratios is to regard them as acceptable if they are greater
than two. This rule is a carry-over from the days when the primary interest
was in verifying if the \(b\)-coefficient was significantly different from zero;
that is, whether the independent and dependent variables were related in
a qualitative sense. But if, as in the models we are considering, the equations
and their coefficients are being used to make forecasts and to prescribe
policy then one would expect that the model builders would look for \(t\)-ratios

\(^1\) The diagonal elements of the symmetric matrix \((R)\) are 1.0; the off-diagonal elements
can vary between -1.0 and 1.0. Thus the value of the determinant \(|R|\) can vary between
1.0 and zero. In the two-variable case if the two regressors are completely independent the
matrix \((R)\) is

\[
\begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix}
\]

so that the determinant is 1.0. If they are perfectly correlated, \((R)\)
is

\[
\begin{bmatrix}
1 & 1 \\
1 & 1
\end{bmatrix}
\]

or

\[
\begin{bmatrix}
1 & -1 \\
-1 & 1
\end{bmatrix}
\]

and so the determinant is zero. As the value of the determinant is
a function solely of the coefficient of correlation between the two regressors it is sufficient
to examine the value of the coefficient. But when more than two regressors are involved
the individual correlation coefficients are not always adequate guides to the value of the
determinant. Consider the following array of four regressors:

\[
\begin{array}{cccccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\
X_1 & 2.0 & 0.0 & 0.0 & 3.5 & 0.0 & 0.0 & 4.0 & 0.0 & 0.0 \\
X_2 & 0.0 & 2.5 & 0.0 & 0.0 & 3.0 & 0.0 & 0.0 & 3.5 & 0.0 \\
X_3 & 0.0 & 0.0 & 3.0 & 0.0 & 0.0 & 2.5 & 0.0 & 0.0 & 4.0 \\
X_4 & 2.0 & 2.5 & 3.0 & 3.5 & 3.0 & 2.5 & 4.0 & 3.5 & 4.0 \\
\end{array}
\]

As \(X_4\) is a linear combination (in this case just the sum) of the other three the value of the
determinant to be used in the regression computations is zero. Yet the individual correlation
coefficients are as follows: \(r_{12} = 0.0; r_{13} = 0.0; r_{14} = 0.59531; r_{23} = 0.0; r_{24} = 0.54972;\)
and \(r_{34} = 0.58600.\) Compared with the correlation coefficients one encounters in the models
under consideration none of these is alarmingly high. In fact, as there are only nine observa-
tions one could even maintain that none of the coefficients is significantly different from
zero at, say, the 2.5 per cent significance level.
that are much higher than two. A t-ratio of two, after all, means that the
standard error of the coefficient is one-half as large as the coefficient itself.
Consequently, the forecasts and the policy conclusions one can draw from
it will indeed cover a very wide band. Second, the fact that the standard
errors increase with increasing collinearity has been taken to imply that
one need not worry about multicollinearity as long as the standard errors
do not increase so much as to render the coefficients insignificantly different
from zero; as long, say, as the t-ratios are not below the magic figure of two.
This is a dangerous deduction. The fact that the standard errors are not
unacceptably high may just be a fortuitous result of the computation
algorithm one has used or of some special inter-relationships among the
variables.

Consider an equation with two regressors: \( Y = b_1 X_1 + b_2 X_2 + u \). The
variance for \( b_1 \) is

\[
S_{b_1}^2 = \sum y^2 (1 - R_{y.13}^2)/(n-k-1) \sum x_i^2 (1 - r_{12}^2),
\]

where \( y^2 = (Y - \bar{Y})^2, \quad x_i = (X_i - \bar{X}_i), \quad R_{y.13}^2 \) is the coefficient of multiple determi-
nation and \( r_{12} \) is the coefficient of correlation between \( X_1 \) and \( X_2 \). The
greater the linear dependence between \( X_1 \) and \( X_2 \)---i.e., the closer \( r_{12} \) is to
1.0---the closer the entire denominator is to zero and, hence, given other
things, higher is the standard error. But in a very large proportion of
multiple regression equations reported, for instance, in UNCTAD, 1968
while the intercorrelation among regressors is 0.95 or 0.9985 the \( R^2 \) term
in the numerator term is also close to 1.0. Thus the term \((1 - R_{y.13}^2)/(1 - r_{12}^2)\)
may not be very large, the effects of extremely high collinearity being
'absorbed', 'submerged', or 'neutralized' by the equally high \( R_{y.13}^2 \).

Technically the estimated coefficients are unbiased and among the class
of linear estimators they are the best. Moreover, as the relatively low
standard errors testify, they are also relatively 'precise'. A coefficient may

\[ \text{1 The expression used in the text can be derived as follows in the two variable case.} \]
\[ \text{We know that} \]
\[ \begin{bmatrix} \sum x_1^2 & \sum x_1 x_2 \\ \sum x_1 x_2 & \sum x_2^2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \sum x_1 y \\ \sum x_2 y \end{bmatrix}. \]
\[ \text{where} \]
\[ y = (Y - \bar{Y}), \quad x_1 = (X_1 - \bar{X}_1) \quad \text{and} \quad x_2 = (X_2 - \bar{X}_2). \]
\[ \text{Thus} \]
\[ \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \sum x_1^2 & \sum x_1 x_2 \\ \sum x_1 x_2 & \sum x_2^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum x_1 y \\ \sum x_2 y \end{bmatrix}. \]
\[ \text{The inverse matrix is:} \]
\[ \frac{1}{(\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2)} \begin{bmatrix} \sum x_2^2 & -\sum x_1 x_2 \\ -\sum x_1 x_2 & \sum x_1^2 \end{bmatrix}. \]
\[ \text{The variance for} \ b_1 \text{ is} \]
\[ S_{b_1}^2 = \text{MSE}[\sum x_1^2/(\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2)], \]
\[ \text{where MSE is the Mean Square Error.} \]
\[ \text{Thus} \]
\[ S_{b_1}^2 = \text{MSE}[1/\sum x_1^2(1 - (\sum x_1 x_1)^2)/(\sum x_1^2 \sum x_2^2))] \]
\[ = \text{MSE}[1/\sum x_1^2(1 - r_{12}^2)] \]
\[ = \sum y^2/(1 - R_{y.13}^2)/(n - k - 1) \sum x_i^2 (1 - r_{12}^2)]. \]
\[ \text{2 Compare Johnston, op. cit., pp. 204–6.} \]
have all these properties and yet may be unreliable for policy purposes. A sample of observations, even when it yields the best linear unbiased estimates of coefficients that have high t-ratios, may yet not be rich enough to be able to disentangle the separate effects of the regressors in an economically meaningful sense.

A set of equations reported in *UNCTAD, 1968* provide an instructive illustration. Faaland and Dahl fit the following regression to account for the value of agricultural output in Tanzania:¹

\[ P_A = 34.322 + 0.402 \text{ (GDP)} \quad R^2 = 0.94, \quad DW = 1.59, \]

\[(0.036)\]

where \( P_A \) is the value of agriculture and the expression in parentheses is the standard error of the coefficient. To take account of the fact that demand patterns may be changing because of the growth of *per capita* income and the diminishing importance of the subsistence sector they fit another regression:

\[ P_A = -8.001 + 5.698 \text{ (GDP/N)} \quad R^2 = 0.96, \quad DW = 1.57. \]

\[(0.397)\]

Then, to see how the relationship between agricultural output and GDP is being modified by changing income levels, they fit an equation containing both GDP and GDP per head as explanatory variables:

\[ P_A = -131.716 - 1.226 \text{ (GDP)} + 22.816 \text{ (GDP/N)} \]

\[(0.239) \quad (3.345) \quad R^2 = 0.99, \quad DW = 2.13. \]

The coefficient of correlation between (GDP) and (GDP/N) is 0.9985. The authors point out that the ‘negative partial correlation (of \( P_A \)) with GDP ... on *a priori* grounds appears improbable’. Yet the t-ratios are high and, if we assume that the third equation is correctly specified, the estimated coefficients are the best linear unbiased estimates.

When collinearities among regressors and the numerical algorithms used for computation lead to patently implausible results one is at once made aware of their distorting influence. But this influence may be present even when the resulting coefficients do not appear to be obviously or dramatically unreasonable. After all the negative and significant coefficient for (GDP) in the third equation cited above is only an extreme illustration. A tightly reasoned and precisely formulated hypothesis can be a great help in such situations. In practice, however, too little attention is paid to formulating precise hypotheses before sending the data down to the computers. As the initial hypotheses are not precisely formulated often

¹ *UNCTAD, 1968*, p. 204.
there is little reason to scrutinize or doubt the results obtained. Consequently, the text accompanying the equations just sets out either to rationalize the values that emerge from the calculations or to reject the results in a perfunctory manner.

From current practice it would seem that for many relationships it is possible to rationalize a wide range of values of coefficients including positive as well as negative signs. Examples can be cited about the direction and extent of the relationship between inflation and consumption, the determinants of Government consumption expenditures, changes in consumption expenditures and their impact on investments via their effect on expectations and so on. A single example may suffice. Consider the relationship of the existing capital stock to current or future gross investment. A negative sign or a very low value of the coefficient could reflect the depressing effects of a larger existing stock on net investment—a hypothesis that is not altogether unreasonable in the kind of capital-stock-adjustment or accelerator models in which demand considerations dominate. On the other hand, a positive and numerically significant coefficient could reflect, first, the stimulus that a more adequate infrastructure provides for investment and, second, the purely arithmetical fact that a larger capital stock requires larger replacement outlays. As there seldom are reliable estimates for replacement outlays and, very often, not even sectoral investment detail, one has to make do with treatments such as the following:

The role of the capital stock in an investment function has been of concern to many researchers. Various alternative specifications of the function have been tried and the coefficient invariably bears a positive sign. Others working with developed economies have found it to be negative. Of course, in a developing economy the positive sign may be valid. However, the capital stock variable has not been included in the relevant equation in the model.1

The scope for rationalizations is even greater when what is at issue is not the sign but the numerical magnitude of a coefficient. Ought we to expect constant returns to scale when fitting aggregate production functions for developing economies? What is a reasonable income elasticity of demand for imported consumer durables? ‘Common sense’ or one’s ‘feel for figures’ is a poor guide on these issues: its only criterion is that the result under consideration should not look too different from past results. And in many cases the past results in their turn are only based on a more or less unsystematic examination (for instance, with the aid of some earlier inadequately specified regression model) of more or less unsatisfactory data.

We see, therefore, that a very large proportion of the regressors that are used are highly collinear, that the resulting coefficients may have little

economic meaning and that we may not always be able to detect this either by using statistical tests like the t-test or by appealing to our intuition. Given these facts, how do the model builders tackle the problems posed by collinear regressors? The method that is most often mentioned in econometrics texts as a way of dealing with collinearities is to use some external or a priori information to supply estimates for the coefficients of all but one of the collinear variables. In practice this ‘solution’ is seldom adopted—reflecting no doubt the difficulties of obtaining a priori estimates for the coefficients. Nor do the models extract the principal components of the regressors and use these instead of the regressors in the equations. Perhaps the reason is that the model builders are interested in the regressors themselves rather than in entities called their ‘principal components’ which they cannot very readily manipulate for policy or projection purposes. The ‘solution’ that is most frequently adopted in practice is to drop one or more of the collinear regressors from the estimating equation. Once again a number of examples could be cited, of which two may suffice.

So as to find a function determining Government consumption expenditures, de Vries and Liu try a number of explanatory variables and report three equations:¹

\[ C_G = -337.8 + 8.96N \quad R^2 = 0.95, \quad d = 1.34 \]
\[ = 31.6 + 0.62T \quad R^2 = 0.94, \quad d = 1.42 \]
\[ = -229.0 - 0.30T + 48.0N - 61.7t - 5.87P_3 \quad R^2 = 0.97, \quad d = 1.34 \]

where \( N \) is population, \( T \) is taxes, \( t \) is time, and \( P_3 \) is a price index. The entries in parentheses are t-ratios. The authors choose the first equation ‘because it seems more important in expressing the economic reason behind an increase in government consumption’. The third function is rejected because ‘though showing significant parameters for all four variables, the sign of \( T \) is reverse to the expectation and others are due to multicollinearity’.¹ Notice the extent to which the coefficient for \( N \) in the equation that is selected (8.96) differs from that in the third (48.0).

On six occasions authors in *UNCTAD, 1968*² postulate that output in the agricultural sector and in ‘other sectors, mainly manufactures’ is a function of total GDP and of GDP per capita. When the output of either of the sectors is regressed on GDP alone or on GDP per capita alone it yields highly ‘significant’ coefficients. When the two regressors are used together the values of the coefficients and their standard errors alter dramatically.

¹ B. A. de Vries and J. C. Liu, op. cit., p. 11.
Some of the coefficients are no longer 'significantly different from zero'.
 Though some are—as in the equation for Tanzania’s agricultural output cited above. But in each of the six cases after a perfunctory discussion of the standard errors being 'relatively high' or of a sign being 'improbable' on a priori grounds the second regressor—per capita GDP—is dropped.

The authors seldom pause to notice that by dropping a regressor or a group of regressors to avoid problems of collinearity they are altering the reduced form of their models very considerably and hence the policy conclusions that can be drawn from them. Moreover, and this leads us to the next sub-section, by dropping these variables they introduce serious specification errors into their equations.

(b) Specification errors: the choice of explanatory variables

The second and in practice an equally important group of reasons for which it may be misleading to treat individual coefficients as the partial derivatives of calculus has to do with the set of explanatory variables that have not been included in an equation. The standard argument for least squares proceeds as follows: let \( X \) and \( Z \) be matrices and \( Y, b, \beta, \) and \( \theta \) be vectors; assume that the true model is

\[
Y = X\beta + u. \tag{i}
\]

Least squares procedures estimate \( \beta \) by \( b \) where

\[
b = (X'X)^{-1}X'Y. \tag{ii}
\]

Substituting for \( Y \) from (i) and taking expectations we get

\[
E(b) = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'E(u). \tag{iii}
\]

But as \( (X'X)^{-1}X'X = I \) and \( E(u) = 0 \)

\[
E(b) = \beta. \tag{iv}
\]

That is, \( b \) is an unbiased estimator of \( \beta \). Suppose, however, that the correct model is

\[
Y = X\beta + Z\theta + w \tag{v}
\]

while the model used for estimation is

\[
Y = X\beta + v. \tag{vi}
\]

That is, let us assume that the analyst has left out some of the explanatory variables because, as we saw in the preceding section, he wants to avoid problems arising from collinearities among regressors, because he has misconstrued the causal relationships, because he cannot obtain satisfactory data about them, or for some other reason; then we obtain biased estimates of \( \beta \). As before,

\[
b = (X'X)^{-1}X'Y. \tag{vii}
\]
Substituting for $Y$—but this time from (v)—and taking expectations we get

$$E(b) = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'Z\theta + (X'X)^{-1}X'E(w).$$ (vi)

As $(X'X)^{-1}X'X = I$ and $E(w) = 0$ we have

$$E(b) = \beta + A\theta,$$ (vii)

where $A$ is $(X'X)^{-1}X'Z$ and is known as the alias matrix. Thus the omission of a variable that has some causal links with $Y$ means not only that in the end we do not have an estimate of the coefficient linking the excluded variable with $Y$ but also that the estimates of coefficients linking all the other $X_i$ to $Y$ may be vitiated. Incidentally, if the $Z$ are included when in fact $Y$ is independent of them, or if they are left out when (though influencing $Y$) they are linearly independent of $X$, then the estimates of $\beta$ are not vitiated. If $\theta = 0$ or if $A = 0$, then $E(b) = \beta$ in (vii).

If we are interested only in forecasting and have reason to believe that the future relationship between $(X)$ and $(Z)$ will be the same as in the past, there is nothing wrong in using an abbreviated model $[Y = f(X)]$ and having the $(X)$ do the work for themselves as well as for the $(Z)$. But if we are going to use the coefficients for policy prescription then this is precisely the situation when leaving out $(Z)$ as in the abbreviated model will give us highly misleading estimates of the coefficients linking $(Y)$ and $(X)$. Instead of leaving $(Z)$ out, the model-builder should search out the parent variables that explain $(Z)$ and $(X)$ themselves—if such variables exist—and substitute those variables into the equation. If the search does not yield the ultimate explanation for the collinearity of $(X)$ and $(Z)$ one can indeed argue that both sets of variables, $(X)$ and $(Z)$, should be included even if the $t$-ratios of one or both are not significant so that mechanical regression procedures do not attribute to one of them the variance of the dependent variable that the two together help explain. From this negative point of view a variable may be very important even if it is ‘statistically insignificant’.

We have already stressed that many econometric models are being used for control purposes also. The question that arises then is: how often is it that in practice explanatory variables that have a quantitatively significant influence on the dependent variables get left out from one’s estimating equations? One has only to scan the text (and perhaps the footnotes) accompanying the equations of a model to realize that—for one reason or another—every author has had to leave out quite a few variables from most equations that he would have liked to include.\(^1\) Apart from the fact that

\(^1\) One repeatedly comes across passages like the following: ‘The suitable explanatory variable for raw material imports seems to be the level of output in secondary industry (manufacturing, construction, and electricity). However, in the absence of necessary data, the following two alternative forms (linking imports of raw materials to fixed investment and to GDP) were obtained’, *UNCTAD, 1968*, p. 360. ‘The model fails to include all the relevant variables affecting inflation and growth either because of the lack of statistical
for many of the important explanatory variables the relevant data is not available, and that some of them are very difficult to quantify anyhow, there is the pervasive difficulty that economic theory offers insufficient guidance as to what constitutes an adequate list of explanatory variables. Some models assume very simple functional relationships; others are more venturesome and imaginative, making the birth-rate a function of per capita consumer expenditure and time; rates of exchange a function of capital transfers into the country in the preceding period, the ratio of the official exchange rate to the GDP deflator and the ‘capacity to import’ which in turn is to be a function of the price index of the country’s imports; and investment a function of earnings of the country from tourism and net credit extended to the private sector.

Now, there is no logical reason why any of these or more esoteric variables should not have been included; but the list does suggest that when one opts for one or a very small number of explanatory variables one has possibly not exhausted the field for potentially relevant candidates. Thus the risk is always present that the estimates of the coefficients are misleading. In a sense, the fact that in devising a model the model-builder almost invariably has a specific purpose in mind (for instance, to project the two gaps, to analyse the consequences of different exchange-rate policies, to estimate the ‘trade-offs’ between various rates of growth and inflation) itself greatly influences his choice of explanatory variables. This becomes apparent when three or four models for the same economy are compared. The sets of equations in them differ much more in their list of explanatory variables than they do in their lists of dependent variables: a model aimed at analysing the choices between inflation and growth attempts to include price variables at as many points as possible; another shows a marked preference for the exchange rate; a third for ratios depicting structural changes in the economy. As the purpose influences the selection of explanatory variables, it may indirectly influence the values and signs of the coefficients derived and, thus, the numerical conclusions that are drawn from the model.

Each of the structural equations in a model can be specified in a variety of ways. When we use the model as a whole—for example, focusing on the import–export gap (a magnitude that depends on a large number of equations)—the uncertain specifications of each structural equation are communicated to the final result; we cannot be certain that the biases will cancel and not reinforce each other. This is especially important because...
often, as in the case of the import–export gap, we are interested in what are relatively small differences between fairly large variables so that even small errors in the latter are liable to affect the former in a major way; in fact, there often are other variables in the system—a country's debt-servicing obligations are a case in point—that depend on the cumulative values of these small differences. Moreover, just as there is no unique way of specifying a single equation so is there no unique way of specifying the model as a whole. For example, the conventional two-gap models are designed to project the investment—saving gap and the import–export gap independently and are by their very nature overdetermined. There are many ways of reconciling the two gaps and the outcome in terms of the final projections and policy prescriptions will differ depending on which particular reconciliation procedure is adopted. Considering the numerous possibilities open to him in specifying a model and its individual relationships the model-builder may find it very difficult to be sure that his projections and policy prescriptions are anything more than a mechanical result of the particular way in which he has specified his model.

IV. Functional form of the equations

One way in which a model-builder may mis-specify the relationship between a dependent variable and a set of independent variables is to assume that the relationship is, say, linear when in fact it is, say, quadratic. This is a particular case of the more general class of specification errors considered above; but a few additional comments are needed.

The choice between one functional form and another is not a matter of indifference. First, each form implies a very specific view of underlying economic relationships. There is nothing in economic theory or in the empirical research conducted thus far which tells us that the most reasonable assumption about the future is that the marginal rate will remain unchanged (the assumption one makes when one uses a linear equation of the $Y = a+bX$ variety). It may be just as reasonable to assume that both the average and marginal rates will remain constant and use the relationship $(Y = bX)$; or to assume that both will vary but that the elasticity will remain unchanged and use $(\log Y = \log a+b \log X)$; or to assume that the elasticity will fall as $Y$ rises and use a semi-log relationship $(Y = a+b \log X)$; or that the elasticity will fall as $X$ increases using, then, the log-inverse function $(\log Y = a-b/X)$; or that the average and marginal rates as well as the elasticity will vary without any particular restriction as, say, in the quadratic equation $(Y = a+bX+cX^2)$. In dealing with a few problems one can invoke some consistency conditions (for instance the homogeneity conditions in demand analysis) to limit the choices open to the model-builder or to constrain the coefficients in some particular way. But, for
most of the relationships that highly aggregated models of the kind con-
considered here try to quantify, our knowledge of economic relationships is
just not specific enough to suggest unambiguously that we should use one
form and not another.
Second, one cannot just abdicate responsibility and choose that particular
form which gives 'the best fit' to the sample period data. For in many
cases various alternative forms yield similar \( R^2 \)'s. But, it may be asked,
if different forms fit the sample period data equally well, does it make much
practical difference which is used for making projections or drawing policy
conclusions? Indeed it does. Consider the following equations which have
been computed from Kenyan data for 1960–8.

\[
\begin{align*}
(DIR\text{TAX})_t & = 0.04847 (GDP\text{FAC})_t \\
 & (0.00191) \quad (F\text{-ratio}) \quad R^2 = 36.997 \quad 0.822
\end{align*}
\]

\[
\begin{align*}
(DIR\text{TAX})_t & = -9.03077 + 0.07468 (GDP\text{FAC})_t \\
 & (0.39871) (0.00704) \quad 112.502 \quad 0.933
\end{align*}
\]

\[
\begin{align*}
\log(DIR\text{TAX})_t & = -2.54335 + 1.48174 \log (GDP\text{FAC})_t \\
 & (0.00969) (0.13391) \quad 122.441 \quad 0.938
\end{align*}
\]

\[
\begin{align*}
(DIR\text{TAX})_t & = 34.15114 - 0.18211 (GDP\text{FAC})_t + \\
 & (0.19761) (0.05425) + 0.00037 (GDP\text{FAC})^2 \\
 & (0.0008) \quad 246.242 \quad 0.984
\end{align*}
\]

\[
\begin{align*}
(M\text{CAPTL})_t & = 0.27497 (INVEST)_t \\
 & (0.01945) \quad 34.020 \quad 0.614
\end{align*}
\]

\[
\begin{align*}
(M\text{CAPTL})_t & = -5.28221 + 0.36031 (INVEST)_t \\
 & (1.02273) (0.05305) \quad 46.125 \quad 0.849
\end{align*}
\]

\[
\begin{align*}
\log(M\text{CAPTL})_t & = -1.37567 + 1.44553 \log (INVEST)_t \\
 & (0.03465) (0.25389) \quad 32.417 \quad 0.797
\end{align*}
\]

\[
\begin{align*}
(M\text{CAPTL})_t & = -27.87714 + 1.14340 (INVEST)_t - \\
 & (0.96160) (0.56760) - 0.00604 (INVEST)^2 \\
 & (0.00436) \quad 27.047 \quad 0.867
\end{align*}
\]

\[
\begin{align*}
\log(M\text{CAPTL})_t & = 1.74490 - 28.71198 [1/(INVEST)_t] \\
 & (0.05169) (2.49690) \quad 132.228 \quad 0.943
\end{align*}
\]

In the first set of four equations receipts from direct taxes are made a
function of GDP at factor cost and in the second set of five equations
imports of capital goods are made a function of gross fixed investment.
Figures in parentheses are standard errors. Each of the equations yields
reasonably good fits for the sample period and the \( t \)-ratios are satisfactory
for all of them. In Figs. 1 and 2 the different equations are used to project
the dependent variables into the future. In each figure identical values
of the independent variable are used for all the equations. The diagrams
show very clearly that the different forms yield widely divergent projections
for the future.
Fig. 1 and Fig. 2. Effects of functional form on projections. Actual values in the sample period are represented by (●). Values of $R^2$ are given in parentheses.
The same sort of differences arise when one uses the first derivatives of the reduced form equations to assess the consequences of a change in some of the independent variables. While the first derivatives of equations like \( Y = aX \) or \( Y = a + bX \) are constant and unambiguous, those of equations like \( Y = aX^2 \) or \( Y = a + bX + cX^2 \) vary with the value of \( X \). Therefore, their absolute values and, in many cases, the rankings of, say, individual industries according to their effect on the import–export gap, may be very different depending on whether one takes \( X \) to be equal to the mean of the sample period or assigns to it the value it is projected to have at the end of the projection period. As the values of the derivatives vary, so must one’s policy prescription.\(^1\)

We do not have to seek far for the reasons that account for these differences. Least squares estimates of coefficients are fairly sensitive to extreme values. As the sample consists of only five to fifteen observations even a few aberrant values can affect the estimates very considerably. Moreover, as the equations are used to project variables many time periods beyond the sample period or to gauge the impact of very substantial changes in the ‘policy variables’, even small differences in the marginal rates of change implicit in the equations get magnified in a very substantial manner.

Finally, the choice of a particular functional form is not without its consequences for the efficacy of initial estimates of the coefficients. It can be shown—and the case is just a particular variety of specification errors considered above—that in general a linear equation will not yield an unbiased estimate of the true marginal rate even close to the centre of gravity of the observed point set when the true underlying relationship is, say, quadratic or log-linear. Biases arise both because of non-linearity in the true relationship and because of an erroneous assessment of disturbances.\(^2\)

V. Mechanical use of certain statistical tests

The question now comes to mind: do not standard statistical tests alert us to the sorts of deficiencies that have been illustrated above? For example, if some important explanatory variables have been left out of an equation can it not be assumed that the value of the coefficient of determination will be extremely low or that the residuals will be highly auto-correlated

---

\(^1\) If the data has some non-linearities in it non-linear equations like the quadratic one are bound to explode when extended far enough into the future. The point being made in the text is not that quadratic equations should always be used for long-term projections even when they yield patently absurd results; rather that the data often contain significant accelerations or decelerations, and if so these should be examined instead of being mechanically suppressed in a linear equation.

and that the standard tests will draw our attention to the excluded variables? Similarly, if the independent variables are highly collinear will the low values of the t-ratios not warn us about their consequences? These questions need to be considered at some length; the answers unfortunately are far from reassuring.

Most models display statistics relevant to three statistical tests: the coefficient of determination, the t-test for individual coefficients, and the Durbin–Watson test for auto-correlation in the residuals. Among these the coefficient of determination receives the greatest attention. In fact, many models display an almost obsessive concern with the $R^2$.

Consider some extracts from a paper projecting employment patterns over twenty years:

‘Group 0. For professional workers, the semi-logarithmic form gave the best overall fit and appeared to cope with elements of non-linearity in the relationship. The equation chosen was:

$$S_0 = -12.05 + 2.94 \ln Y + e,$$

(0.21)

with $R^2$ equal to 0.843. Only per capita income shows a regression coefficient significant at the five percent level, although the linear form (with smaller $R^2$ of 0.811) shows population size significant as well . . .

‘Group 2. A semi-logarithmic form was shown to fit the data on clerical workers quite well ($R^2 = 0.784$), with no evidence against the linearity of this relationship. Although growth rate as well as income was significant in the simple linear fit, $R^2$ was lower (0.766). Hence, the equation chosen was:

$$S_2 = -13.12 + 3.22 \ln Y + e.$$

(0.28)

‘Group 4. For farmers and related occupations, the best fit is found with the semi-logarithmic relationship:

$$S_4 = 153.33 - 19.34 \ln Y + e,$$

(1.40)

with $R^2$ of 0.838. The scatter diagram revealed no problems with the linearity of this fit. The simple linear form shows population and growth rate significant (at the 10 percent confidence level) as well as income, but $R^2$ was much lower at 0.677 . . .

‘Group 9. As with sales workers, neither equation form is very successful in explaining the variance. The semi-logarithmic form, however, had the higher $R^2$ (0.278, significant by F-test), and will be used despite the appearance of non-linear elements in the corresponding scatter diagram.
The equation used is:

\[ S_y = 5.55 + 1.63 \ln Y - 0.78 \ln P + e \ldots \]

Similarly, the Colombia model referred to above cites two equations explaining investment in construction:

\[ \frac{I_{hH}}{P_{cnt}} = -0.9595 + 0.1312X + 1.208 \left( \frac{I_{mex}}{P_{cnt}} \right) \quad R^2 = 0.977, \quad d = 2.327 \]

and

\[ \frac{I_{hH}}{P_{cnt}} = -0.311945 + 0.39590 \left[ X - X_{-1} \right] + 0.5920 \left( \frac{I_{mex}}{P_{cnt}} \right) + \]

\[ + 0.94580 \left( \frac{I_{hH}}{P_{cnt}} \right) \quad R^2 = 0.912, \quad d = 1.590. \]

The coefficients for \( I_{mex}/P_{cnt} \) are very different in the two equations and this should have provided an important clue to the 'reliability' of the coefficients for control. However, the text merely notes that the former equation 'was accepted in the final analysis on the basis of its predictive value'.

In a superficial sense this preoccupation with the \( R^2 \) is well rewarded. The models are fairly simple and highly aggregated representations of the economies of developing countries. Few of them incorporate any institutional detail or any information about production relationships in the economy. Yet if one considers their structural equations individually one is struck by their very impressive coefficients of determination. It is often and rightly said that the process of growth and change in developing countries is a complex and in many respects an incompletely understood phenomenon. The equations of these models, on the other hand, seem to suggest that the broad features of growth and change in a host of dissimilar countries over the past five to fifteen years can be represented by extremely

2 K. Marwah, op. cit., pp. 233 and 236–7. \( (X) \) is GNP at 1958 prices; \( I_{mex} \) is value of imports of construction materials, and \( P_{mex} \) and \( P_{cnt} \) are indices of prices of construction materials in pesos and dollars.
3 The eighteen models in UNCTAD, 1968 utilize about 370 estimated equations (though, as noted elsewhere, about 630 estimated equations are reported in the text). Their distribution by the number of regressors used in the equation is as follows.

<table>
<thead>
<tr>
<th>Total number of equations</th>
<th>Equations with 1 regressor</th>
<th>Equations with 2 regressors</th>
<th>Equations with 3 regressors</th>
<th>Equations with 4 or more regressors</th>
</tr>
</thead>
<tbody>
<tr>
<td>370</td>
<td>281</td>
<td>67</td>
<td>21</td>
<td>1</td>
</tr>
</tbody>
</table>
simple equations with no more than one or two, or at the most four or five, explanatory variables.¹

Even though the point has been made that complex social phenomena can often be represented by fairly simple mathematical expressions,² the high degree of success of the models and their simple equations, in that they seem to explain a wide range of dependent variables, comes as a considerable surprise. However, this success of the models in explaining developments in particular economics is almost entirely deceptive. And the models pay a substantial price for this apparent success. The foregoing extracts illustrate one consequence of relying so heavily on the coefficient of determination. By mechanically adopting the equation with the higher $R^2$ the authors in each case exclude a number of explanatory variables from the particular segment of their model. The reduced form and, perhaps, also the dynamics of the model and hence the policy conclusions that can be derived from it are all altered drastically. Now, it so happens that the coefficients of determination can often be affected by so many extraneous factors, which may have nothing to do with functional or causal relationships between variables, that their use can be very misleading—especially when the models or equations are to be used as a basis for policy prescription. Three sets of considerations will be used to illustrate some common situations in which reliance on the $R^2$ can mislead the model-builder.

By far the most frequently encountered situation is the one that Yule documented forty-five years ago:³ two series may display ‘nonsense correlations’ because, even though unrelated in a causal sense, they have pronounced trend or cyclical components. Consider, for instance, Kenya's exports and imports of services other than tourism. These consist mainly of freight, transportation, insurance, and similar items. Let us postulate that most of these arise from the movement of goods in foreign trade and, therefore, let us regress them against the single regressor: the total value of goods imported into and exported from Kenya. For different equation forms the $R^2$ for the 1960–8 data for the exports of services are between

<table>
<thead>
<tr>
<th>Coefficient of determination</th>
<th>$&lt; 0.3$</th>
<th>$&lt; 0.4$</th>
<th>$&lt; 0.5$</th>
<th>$&lt; 0.6$</th>
<th>$&lt; 0.7$</th>
<th>$&gt; 0.7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(i)$ UNCTAD, 1968</td>
<td>2.2</td>
<td>1.3</td>
<td>6.1</td>
<td>6.1</td>
<td>11.2</td>
<td>11.7</td>
</tr>
<tr>
<td>$(ii)$ ECAFE, 1968</td>
<td>2.5</td>
<td>5.0</td>
<td>4.2</td>
<td>4.2</td>
<td>11.8</td>
<td>18.5</td>
</tr>
</tbody>
</table>

¹ On a rough count the eighteen models in UNCTAD, 1963 between them use about 370 estimated equations and the eight models of ECAFE, 1968 list about 120 computed equations. The distribution of these equations by their coefficients of determination is as follows:


0.912 and 0.950 and for the imports of services between 0.715 and 0.779. The $F$-ratios are highly significant as are the $t$-ratios. The fact, however, is that Kenya's foreign exchange earnings from 'exports of services' and expenditures on 'imports of services' are related only in an incidental way to its imports and exports of goods: first, Kenya has no large shipping lines in which these goods may be ferried and, second, imports of goods as used in computing the equations are recorded c.i.f. so that whatever impact they have on freight and transportation costs is already taken account of elsewhere. Kenya's earnings from freight and transport arise more from the imports and exports of Uganda and Tanzania rather than its own imports and exports. The incidental way in which its own imports and exports of goods enter the picture is that a small part of Kenya's earnings from exports of services consist of port-charges and part of these are collected from the ships that ferry goods specifically meant for or specifically originating in Kenya. In this case, therefore, the $R^2$ would certainly have misled us if we abandoned our search for explanatory variables once we had hit upon the total value of goods imported into and exported from Kenya. But, it may be argued, what is wrong in using the equations for projections or policy prescription? After all, the variables have moved together in the past. The difficulty is that the less certain one is about the existence of a causal connection between the dependent and the explanatory variables, the less confident one is about asserting that 'as they have moved together in the past it is reasonable to assume that they will move together in the future'. In fact, taking the 'exports of services' as an example, one would be on firmer ground in asserting that these are likely to grow at a rate lower than the growth of the total value of goods imported into and exported from Kenya: Uganda's exports and imports together are expected to grow at a rate about one-half of the rate being currently projected for Kenya's imports and exports of goods and Tanzania has definite plans of developing Dar-es-Salaam which will mean that it will be relying somewhat less on Mombasa in the future. To avoid pitfalls of this kind models of developed economies almost invariably estimate their relationships from data that has been transformed—for instance, by taking its first differences—to avoid the effects of auto-correlation in the series. But hardly any of the newer models for developing countries is estimated in terms of first differences or similarly transformed data.¹

¹ Not one among the eighteen models of UNCTAD, 1968 estimates its relationships using data that has been transformed to remove the effects of auto-correlation in the original series. The argument here is not that models should always be calculated from first differences: in fact, one can think of situations in which taking first differences may itself introduce spurious correlations into the data. The argument simply is that in many of the equations of the aggregative models considered here the high $R^2$ may well be due to factors like trend components in the series. The difficulties outlined in the text are not
The second illustration of how a spuriously high $R^2$ may be obtained relates not to a failure to transform data when it should have been transformed but to transforming it in such a way as to generate once again the kind of situation that Yule warned us against\(^1\)—in which one gets ‘conjunct series with conjunct differences’.

This illustration is provided by the output–investment functions used in *ECAFE, 1968* and *UNCTAD, 1968*.\(^2\) Fourteen of the eighteen models of *UNCTAD, 1968* and seven of the eight models of *ECAFE, 1968* use the following output function:

$$Y_t = f \left( \sum_{i=0}^{t-1} I_t \right),$$

where $Y$ is GDP at factor cost and $I$ is gross fixed investment. The function yields extremely impressive fits: the $R^2$ being almost invariably higher than 0.95. The two series $Y$ and $I$ have pronounced trend components in almost all cases; by cumulating $I$ we are generating a new series which not only has high auto-correlation but the first differences of which are also auto-correlated. The nature of the effect on the coefficients of determination is perhaps best demonstrated by a simple example. Consider three series: $Y_t$, $X_t$, and $Z_t$, where $Z_t = \sum_{i=0}^{t-1} X_t$. $Y_t$ is growing at 10 per cent and $X_t$ fluctuates around a mean of about 11.5.

<table>
<thead>
<tr>
<th>$Y_t$</th>
<th>$X_t$</th>
<th>$Z_t = \sum_{i=0}^{t-1} X_t$</th>
<th>$Y_t$</th>
<th>$X_t$</th>
<th>$Z_t = \sum_{i=0}^{t-1} X_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.10000</td>
<td>12</td>
<td>0</td>
<td>2.35794</td>
<td>10</td>
<td>89</td>
</tr>
<tr>
<td>1.21000</td>
<td>11</td>
<td>12</td>
<td>2.59374</td>
<td>12</td>
<td>99</td>
</tr>
<tr>
<td>1.33100</td>
<td>10</td>
<td>23</td>
<td>2.85311</td>
<td>11</td>
<td>111</td>
</tr>
<tr>
<td>1.46410</td>
<td>14</td>
<td>33</td>
<td>3.13842</td>
<td>13</td>
<td>122</td>
</tr>
<tr>
<td>1.61081</td>
<td>9</td>
<td>47</td>
<td>3.45228</td>
<td>9</td>
<td>135</td>
</tr>
<tr>
<td>1.77156</td>
<td>10</td>
<td>56</td>
<td>3.79749</td>
<td>14</td>
<td>144</td>
</tr>
<tr>
<td>1.94871</td>
<td>12</td>
<td>66</td>
<td>4.17724</td>
<td>8</td>
<td>158</td>
</tr>
<tr>
<td>2.14358</td>
<td>11</td>
<td>78</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Confined to time series models. Just as trends often result in spurious correlations in time series models so may factors like scale mislead us in models based on cross-section data. Liberty is not the only thing that demands eternal vigilance.

\(^1\) G. Udny Yule, op. cit.

\(^2\) The same function is used in Kanta Marwah, op. cit. The function is used for fourteen of the eighteen countries for which *UNCTAD, 1968* contains models. For Argentine and India imports of capital and intermediate goods have been included in addition to cumulative investment. Among the remaining four countries the function was fitted for Kenya, Uganda, and Tanzania also but the regressions revealed a negative correlation between investment and GDP for Kenya and no significant correlation for Uganda and Tanzania and so the authors assumed a capital–output ratio of 3.1 for Kenya and 2.5:1 for the latter two (*UNCTAD, 1968*, pp. 124-5, 205, and 249). For Thailand there is no separate function, the relevant magnitudes being derived from an identity.
In fitting $Y_t$ to $X_t$ we get:

$$Y_t = 3.00572 - 0.06106X_t$$

$R^2 = 0.0125$ $F$-ratio $= 0.165$

The figure in parentheses is the $t$-ratio. But when we cumulate $X_t$ and fit $Y_t$ to $Z_t$ as in the UNCTAD, 1968 and ECAFE, 1968 output functions we get:

$$Y_t = 0.81828 + 0.01933Z_t$$

$R^2 = 0.970$ $F$-ratio $= 430.172$

The very impressive coefficient of determination and the significant $t$-ratio are due solely to the fact that $X_t$ has been cumulated and do not in any way imply a reliable, stable, or significant relationship between $X_t$ and $Y_t$. The reader can indeed experiment with any set of positive random numbers and obtain similar results.

The final illustration of how $R^2$ may be misleading involves regressing one ratio on another. We know that gross fixed investment in a sector $(\text{INVEST})_t$ consists of net investment $(\text{INVNET})_t$ and depreciation $(\text{INVDEP})_t$:

$$(\text{INVEST})_t = (\text{INVNET})_t + (\text{INVDEP})_t.$$  

Assuming that a given proportion $(d)$ of the capital stock is replaced every year and that the capital stock bears a fixed relation to the value added in a sector

$$(\text{INVDEP})_t = d(\text{KSTOCK}) = d \cdot k(\text{GDPFAC})_t = b(\text{GDPFAC})_t.$$  

Therefore

$$(\text{INVEST})_t = (\text{INVNET})_t + b(\text{GDPFAC})_t.$$  

Dividing both sides by $\Delta(\text{GDPFAC})_t$ we get

$$\frac{(\text{INVEST})_t}{\Delta(\text{GDPFAC})_t} = \frac{(\text{INVNET})_t}{\Delta(\text{GDPFAC})_t} + \frac{b(\text{GDPFAC})_t}{\Delta(\text{GDPFAC})_t}$$

(1)

where $k_{gt}$ is the gross capital-output ratio in sector $(i)$, $k_{nt}$ the net capital-output ratio, and $r_t$ the sectoral rate of growth. On fitting equation (1) to the data we obtain the relevant gross capital-output ratios. Given the sectoral growth rates $(r_t)$ we obtain numerical values of the capital-output ratios that can be used for determining, say, the amounts of investment required to assure a certain specified growth of value added in the sectors.

For the sample period we may estimate the $(r_t)$ by fitting the trend equation

$[Y_t = Y_0 (1 + r)^t$ or $\log Y_t = \log Y_0 + t \log (1 + r)]$ to the sample period data.

---

The sectoral growth rates and the gross capital-output ratios estimated from the Kenyan data for 1960–8 along with the coefficients of determination are given below.

<table>
<thead>
<tr>
<th>Sector</th>
<th>(1) Rates of growth in the sample period</th>
<th>(2) $R^2$ for col. 2</th>
<th>(3) Capital-output ratios for the sample period</th>
<th>(4) $R^2$ for equations estimating capital-output ratios</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture, forestry, and fishing</td>
<td>5.39</td>
<td>0.801</td>
<td>1.58</td>
<td>0.979</td>
</tr>
<tr>
<td>Mining and quarrying</td>
<td>1.84</td>
<td>0.000</td>
<td>5.20</td>
<td>0.624</td>
</tr>
<tr>
<td>Manufacturing and repairing</td>
<td>9.27</td>
<td>0.966</td>
<td>2.87</td>
<td>0.987</td>
</tr>
<tr>
<td>Building and construction</td>
<td>3.80</td>
<td>0.134</td>
<td>3.63</td>
<td>0.961</td>
</tr>
<tr>
<td>Electricity and water</td>
<td>3.79</td>
<td>0.966</td>
<td>4.22</td>
<td>0.986</td>
</tr>
<tr>
<td>Transport, storage, and communications</td>
<td>8.84</td>
<td>0.981</td>
<td>4.85</td>
<td>0.751</td>
</tr>
<tr>
<td>Other services</td>
<td>6.86</td>
<td>0.969</td>
<td>2.61</td>
<td>0.970</td>
</tr>
<tr>
<td><strong>TOTAL: When estimated directly</strong></td>
<td><strong>6.56</strong></td>
<td><strong>0.964</strong></td>
<td><strong>2.69</strong></td>
<td><strong>0.999</strong></td>
</tr>
</tbody>
</table>

The coefficients of determination for the trend equations and for the capital-output ratios are satisfactory for all the sectors other than mining and quarrying. In addition, the capital-output ratios do not seem unreasonable. Would we be justified in presuming that the satisfactory $R^2$ and the reasonable values of the ratios indicate that we now have a satisfactory basis for forecasting sectoral gross investment?

Notice that the denominator $(\Delta GDPFAC)_t$ is the same on both sides of equation (1). In fact, it turns out that the variance of $(\Delta GDPFAC)_t$ is much greater than the variance of either $(INVEST)_t$ or $(GDPFAC)_t$. So that the reason we get the very impressive $R^2$ for the capital-output ratios is largely that the variance of $(\Delta GDPFAC)_t$ is just explaining itself. The literature contains formal tests relevant to this situation but they are seldom employed in practice. Whether one is going to regard the very high values of $R^2$ as 'spurious' or not depends to a large extent on one’s predilections; for philosophically the situation is an ambiguous one.


2 The question of spurious correlation quite obviously does not arise when the hypothesis to be tested has initially been formulated in terms of ratios, for instance in problems involving relative prices. Similarly, when a series such as money value of output is divided by a price index to obtain a ‘constant dollar’ estimate of output, no question of spurious correlation need arise. Thus, spurious correlation can only exist when a hypothesis pertains to undeflated variables and the data have been divided through by another series for reasons extraneous to but not in conflict with the hypothesis framed as an exact, i.e. non-stochastic, relation’, E. Kuh and J. Meyer, op. cit., pp. 401–2.
Perhaps the following practical consideration will be more persuasive than lengthy arguments about whether or not the correlations are properly regarded as spurious. The sectoral capital-output ratios reported above were used along with actual year-to-year changes in sectoral (GDPFAC) to estimate fixed investment in the sample period. The following $\chi^2$ result from a comparison of forecasted and actual investment in 1961–8.

Comparisons of actual and forecasted gross fixed investment in Kenya, 1961–8

<table>
<thead>
<tr>
<th>Sector</th>
<th>$\chi^2$: Total 1961–8</th>
<th>$\chi^2$: Years 1961–8 taken individually</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture, livestock, forestry, and fishing</td>
<td>0.002</td>
<td>347.635</td>
</tr>
<tr>
<td>Mining and quarrying</td>
<td>0.037</td>
<td>141.052</td>
</tr>
<tr>
<td>Manufacturing and repairing</td>
<td>0.255</td>
<td>18.071</td>
</tr>
<tr>
<td>Building and construction</td>
<td>0.098</td>
<td>268.240</td>
</tr>
<tr>
<td>Electricity and water</td>
<td>0.026</td>
<td>8.395</td>
</tr>
<tr>
<td>Transport, storage, and communications</td>
<td>0.505</td>
<td>14.011</td>
</tr>
<tr>
<td>Other sectors</td>
<td>0.186</td>
<td>40.502</td>
</tr>
<tr>
<td><strong>All sectors</strong></td>
<td><strong>1.109</strong></td>
<td><strong>837.906</strong></td>
</tr>
</tbody>
</table>

Column 2 indicates that the forecasts of cumulated investment for the period as a whole are indeed satisfactory; the $\chi^2$ for 'all sectors' in column 2 is less than the permissible limit with six degrees of freedom. But the cumulated totals are made up of very large over- and underestimation of investment in individual years; thus almost all the individual $\chi^2$ values in column 3 as well as the $\chi^2$ for 'all sectors' are well beyond their permissible values for seven and fifty-five degrees of freedom respectively. May one conclude that while the capital-output ratios are not reliable for forecasting investment in an individual year such as 1970 or 1971 they are likely to give satisfactory estimates for investments in the 1969–78 period as a whole? Unfortunately, even this much is by no means certain: cumulative investments are close to actual investments only over the sample period taken as a whole; when, for instance, forecasted investments are totalled up for a part of the sample period (say, 1965–8) they are once again vastly different from the totals of actual investments. Thus we cannot be certain that even the cumulated values of investment will be reasonable estimates of the true values for any period other than the sample period taken as a whole. Reliance on the very high $R^2$ of equations estimating capital-output ratios would, therefore, have been completely unwarranted. A part of the reason, of course, is that the $R^2$ of equation (1) above relate to the magnitude
When, in fact, we are interested in—and are implicitly using them to tell us something about—the variable (INVEST),

Besides the coefficient of determination the two tests most frequently used—or at least displayed—are the t-ratio and the Durbin–Watson statistic. Some comments on the t-ratio were included in the discussion of collinear regressors. The Durbin–Watson statistic does not require extended comment.

The importance of the error-term is greatly emphasized in the theory of econometrics, since a critical assumption of regression and econometric models is that the error-term is random. Auto-correlation in the residuals may indicate, for instance, that a systematic element has been left out in the explanation; in addition, least squares estimates are no longer efficient (though they are still unbiased and consistent) and the usual significance tests are no longer directly applicable. The question becomes particularly important in auto-regressive schemes that utilize lagged variables for in such cases the estimates of 'b' are neither unbiased nor consistent. For all these reasons econometric theory places considerable emphasis on a careful examination of the error-term. In this spirit the Durbin–Watson statistic—one of the tests available for examining residuals—is faithfully presented for most equations. But in many cases no further attention is paid to it. Thus, in a particular case, it is stated that 'the (Durbin–Watson) statistic shows that the residuals of the German functions are positively auto-correlated at the five percent level... the subsequent discussion, however, will disregard all complications arising from auto-correlated errors'.¹ Nor is this unrepresentative:² for many of the equations contained in econometric models of developing countries the Durbin–Watson statistic shows significant auto-correlation in the residuals, yet in no case is any attempt made to transform the data or the specification of the equation in any suitable way to avoid the auto-correlation. Of course, it may be that these models do not pay much attention to the statistic because at least thirteen to fifteen observations are needed to be able to use its tabulated values and more often than not the models are based on a smaller number of observations. One suspects, however, that having listed the statistic no one bothers about it simply because no one else bothers about it either: everyone is aware of sinning in good company.

Before concluding this review of the relevance of standard statistical tests, it will perhaps be worth while to stress one general point. The theory of the sampling distribution of $R^2$ and other indices like the t-ratios assumes

² For similar remarks see *UNCTAD, 1968*, pp. 108, 126, 191, 234.
that the hypothesis to be tested has been formulated independently of the data that is used to test its validity. When, as in practice, the hypothesis is framed after a scrupulous and in some cases systematic examination of the data (with, for instance, the aid of step-wise regression procedures) the indices lose a good part of their discriminatory power. To validate the hypothesis or the equation that embodies it one needs data from a different sample—for instance, from a time span different from the sample period used in estimating the equation. As the developing countries seldom have consistent series of more than five to fifteen years one seldom finds a model-builder verifying his equations or hypotheses by reference to data other than the data used in their estimation and formulation.

VI. Point forecasts and policy prescriptions for years far into the future

Even if it were the case that the data from which the equations were estimated was reliable and ample; that the problems arising from collinearities had been successfully taken care of; that we could be sure that the equation had been correctly specified; and if care had been taken to avoid the kinds of traps sketched in Section V—even then the question as to whether the coefficients and equations could be regarded as reliable guides to the future is a separate one, which has to be answered with reference to whatever information one can muster about that future.

Models currently in use make point forecasts for years well beyond the end of their sample of observations. We can be almost certain that the precise value for $Y_T$ we predict from an assumed relationship, say,

$$(Y = a + bX),$$

will never be realized. The individual value of $Y$ is in a sense only an average of likely outcomes. If the forecasts from an equation are to be viewed in the context of some significance or probability levels they must be in terms of a $Y$-distribution and not just a single value of $Y$.

Now, the interval estimate of $Y$ at $(1-\alpha)$ level of confidence that one must work with is given by

$$\hat{Y} \pm t_{(n-p-1),(1-\alpha)} \sqrt{(X'_0(X'X)^{-1}X_0),}$$

where $\hat{Y}$ is the point forecast of $Y$, $n$ the number of observations in the sample period, $p$ the number of explanatory variables, and $X_0$ that point in $X$-space for which the forecasts are being made. The upper and lower boundaries of these interval forecasts curl away from the line of point forecasts the further away $X_0$ is from the means of $X$ in the sample period. Typically, individual equations are being used to make forecasts for periods
when the independent variables are two to five and sometimes an even larger number of times their mean values in the sample period and one and a half to three and sometimes an even larger number of times their values at the end of the sample period. It is not surprising then that when we are conscientious and do make interval projections for the future—even with just one functional form—the upper and lower bounds turn out to be a long way apart from each other.

The point is not merely that one now has a band instead of a single point forecast from an individual (structural or reduced form) equation; these bands have to be combined in some way to yield an interval forecast of, say, the import–export gap to which one may attach a precise probability significance. Unless one has estimated the reduced form equation for the gap directly—in which case there is no need to combine any equations—the attempt to combine conditional forecasts from individual equations soon leads one into problems. For in combining the confidence bands of one structural equation with confidence bands of another one soon loses control of probability levels.

We know that if $R_1, ..., R_n$ are random variables with finite variances $\sigma_1^2, ..., \sigma_n^2$ and $S_n = R_1 + ... + R_n$, then

$$\text{Var}(S_n) = \sum_{k=1}^{n} \sigma_k^2 + 2 \sum_{j<k} \text{Cov}(R_j, R_k)$$

where the last term contains the sum of each of the $\binom{n}{2}$ pairs $R_j, R_k$ with $j < k$ once and only once. In the models considered here the authors—to avoid problems of simultaneity—assume that the residuals of different equations are independent of one another. Therefore, the variance of residuals that one needs in order to obtain a band forecast for, say, the import–export gap, is:

$$\text{Var}(S_n) = \sum_{k=1}^{n} \sigma_k^2,$$

where $k$ refers to the residuals of each of the equations that goes into the reduced form of the gap. For the author’s Kenya model this pooled $\text{Var}(S_n)$ comes to £756.9 million—the standard deviation, that is, is around £27.5 million. As the model projects the gross external gap to be around £20 million to £25 million a year throughout the entire projection period, an interval forecast based on this wide band is as likely to be correct as it is unhelpful.

In the specific context of the conventional two-gap models that are so...

---

1 To cite but one instance, in Jere Behrman and L. R. Klein, *Econometric Growth Models for the Developing Economy*, Wharton School, University of Pennsylvania (mimeographed), 1968, by the end of the projection period some of the independent variables increase to thirty-eight times their values at the end of the sample period.

2 This estimate makes no provision for the many variables in the balance of payments and the debt routine which are exogenously specified in the model.
widely used now the moral of the foregoing discussion might seem to be that if one is primarily interested in forecasting magnitudes like the import–export and the investment–savings gaps one should directly estimate the reduced forms for the two gaps from the sample period data. If one attempts to do so then, clearly, one would have to be even less ambitious while choosing the explanatory variables than the simple models being proposed at present. The number of variables that enter the reduced form for any one of the gaps in these models is often quite large and the number of observations is often so small that there are few degrees of freedom left to estimate the coefficients. Moreover, in attempting to estimate the reduced forms directly there is the immediate difficulty that in the (ex post) sample period data the savings–investment and the import–export gaps will be necessarily identical. Therefore, the only way of obtaining estimates of equations that would yield independent (i.e. not identical) projections of the two gaps for the future will be to specify the model in a way that is very different from the manner in which two-gap models are being specified at present. One may, for instance, fit the reduced forms for the two gaps using different sets of explanatory variables. Alternatively, instead of having different lists of explanatory variables for the two gaps, one could specify the individual equations in the same way as they are specified at present but in addition specify some adjustment procedures that would furnish independent estimates of the two gaps. The difficulty here would be, as was noted towards the end of Section III, that as there are many ways of specifying the equations and reconciling the gaps the forecasts may well turn out to be quite sensitive to the particular specification or reconciliation procedure adopted.

It has been stressed above that these simple models are being used for making projections for periods far beyond their sample periods. Many statisticians who have worked with regressions and seen their forecasts go astray will express serious doubts as to whether regression equations can be used at all for values of variables that are two to five times (to say nothing of variables that are thirty-eight times) beyond the range of values covered by the sample observations. Using these equations implies the heroic assumption that the interrelationships of the system in the future will be as they have been in the past or that structural change in the future will follow the pattern of whatever change occurred in the sample period when, in fact, the whole effort of the governments of these countries and others connected with their development is to bring about a structural transformation in the economies. Nor is it the case that when information about likely structural changes becomes available it is customarily incorporated in the model. The fact is that in most of the models equations are being used in a fairly mechanical manner for projections five or ten years
into the future. But even if the analyst using a very aggregative econometric model sought out and obtained specific information about a likely structural change in, say, the manufacturing sector, how would he incorporate it into his aggregated model? The models in use are at such a level of aggregation that if one were told, for instance, that in Kenya canning of fruit and vegetables and leather and fur processing are going to grow at twice, sugar, spinning, and weaving and pulp and paper at about two and a half times, and rubber at four times the rate of growth of the manufacturing sector as a whole or that Kenya is going in for a large fertilizer plant using by-products of its refinery as feedstock one would not be able to make full use of this kind of specific information. All one could do would be to lump it all together to see what effect it has on the overall growth rate of the manufacturing sector or of the economy and work with that. If by contrast the model had concentrated on incorporating a greater amount of production detail it would have been able to utilize this kind of specific information effectively. And it need hardly be stressed that information about the future when it becomes available almost invariably comes in this specific form; seldom is there a revelation about the overall marginal elasticity of consumer imports or the overall capital-output ratio in a sector.

The position is thus as follows: equations are being used to make forecasts for periods well beyond the range of observation points covered in the sample period; the only information usually obtainable about those distant periods relates to specific and particular developments; and the models and equations are at such a level of aggregation that they cannot use this information. Interval forecasts, though far preferable to point forecasts, do not help partly because of the practical difficulties outlined above and partly because the structural change that makes point forecasts undependable also makes the information on which interval estimates are based obsolete. For both sets of forecasts are based on the same information.

VII. The moral of the story

The argument of this paper has not been against quantification or sophistication in economic analysis nor against economic models in general. It has focused on a particular type of model that is being used for analysis in developing countries: the highly aggregated macro-economic regression model. Over the last few years models of this kind have become increasingly common and are being used for medium- and long-term forecasts and for policy prescription. The apparent faith in these models is unwarranted.

As the sorts of models we have been considering are being constructed for one country at a time they necessarily have to rely on time series data. This data is generally available for only a few years, and is of uncertain or very low reliability. As the time series are very short and as almost all
variables seem to grow or decay over the short periods for which data about them can be assembled, every attempt to include in the equations all the independent variables that appear to be relevant on a priori grounds soon exhausts the degrees of freedom and is itself frustrated by the prevalence of collinearities. For this and other reasons there is every likelihood that the models and their equations are grossly mis-specified. Moreover, in a large number of cases one can expect little help from economic theory, from mechanically applied statistical tests, or from appeals to intuition.

Any one of these obvious objections taken individually is sufficient to dissuade one from taking these macro-models at all seriously. In combination their effect can fairly be described as devastating.

It is often said that it is safe to use these grossly under-specified equations and models provided one is aware of their limitations. This is no more than a ritualistic incantation. For in practice one either uses a model, an equation, or a coefficient, or one does not. How is awareness of their limitations usable in a concrete way when a model often involves a number of equations, when little is known about the errors in the data, and when little has been done to experiment with alternative specifications?

Before constructing an econometric model and perhaps again after having constructed it, one should regain one's perspective by rereading some of the iconoclastic literature on regression and econometric analysis (for example writings of Yule, Box, Geary, and others) and on problems presented by the inaccuracy of basic data. National accounts and the manner in which they are put together in a country deserve more attention than they usually receive at present. A plea should also be made for paying greater attention to textbooks that so often warn us against doing so many of the things we do. A classroom is not their only proper place.

International Bank for Reconstruction and Development, Washington, D.C.

1 'The drawbacks to the use of capital–output ratios are well known. . . . Despite these drawbacks, the capital–output ratio is often heavily relied on as a tool of analysis. If enough care is exercised in its application, and particularly if separate estimates can be made for individual sectors, it should be possible to indicate the broad order of magnitude of investment requirements for particular rates of growth', UNCTAD, 1968, p. 13.

