Endogenising Demographic Variables in Demo-Economic Models: The Bachue Experience

RENE WERY and GARRY RODGERS*

This paper surveys the problems encountered and solutions adopted in incorporating demographic phenomena in the main Bachue case studies undertaken to date, namely the demo-economic models constructed for the Philippines, Kenya, Brazil and Yugoslavia. Eight issues are treated: population accounting and lag structure; fertility; mortality; migration; nuptiality; household formation; schooling; and labour force participation. In each case there is a discussion of model structure, dependent and explanatory variables, and empirical strategy. Summary tables compare the approaches of the different models. Although the objective is not to identify the best solution — it is noted that the specifics of every country situation rule this out — some suggestions about more promising approaches are made with respect to choice of variables to include, and the estimation of behavioural models.

I. INTRODUCTION

The purpose of this paper is to describe, discuss, and draw lessons from the treatment of behavioural demographic variables in the Bachue models. These variables are endogenous elements of a wider demo-economic system containing other components which are not described here. In particular, we do not discuss the endogenisation of those non-demographic variables which affect demographic behaviour. We instead concentrate on certain theoretical, technical and practical problems encountered in inserting demographic variables in the system as a whole; how they have been measured in the various applications of the Bachue models, how they are behaviourally explained and linked to the other elements of the system, the data sources used, and some issues of econometric estimates and modelling. The models considered here are those for the Philippines, described in Rodgers, Hopkins and Wery [39], Kenya, a technical description of which is given in Anker and Knowles [4], Brazil, a summary description of which is available in Braganca, Burle de Figueiredo and Rato [8; also 15], and Yugoslavia, described in outline in Macura, Popovic and Rasevic [25]. A general description of all these models is given in ILO [19].

*The authors are associated with the ILO, Geneva.
One objective of the paper is to draw from the Bachue applications general principles on the modelling of demographic variables in a demo-economic system. However, the diversity of the various models is such that generalisations are sometimes difficult to attain. Therefore in presenting the experience with different Bachue models, the stress is on illustrating and comparing the solving of particular problems, identifying difficulties, and discussing the advantages and disadvantages of particular paths to modelling objectives, rather than pursuing the mirage of a universally valid approach.

The demographic variables treated endogenously in the Bachue models are:
(i) fertility,
(ii) mortality,
(iii) migration,
(iv) nuptiality, and
(v) household formation.

Two "para-demographic" elements are also considered:
(vi) schooling and
(vii) labour force participation.

Fertility and mortality are the two basic components of population growth; migration is the third basic behavioural variable required for tracking a spatially disaggregated population. The four other variables have slightly different functions. They constitute criteria to further disaggregate the population according to categories relevant to the network of relationships between demographic and economic phenomena. Whereas the latter are fundamentally quantitative, schooling, labour supply and so on have important qualitative aspects, permitting for instance a distinction between illiterates and literates, or between "working" and "non-working" populations. These qualitative characteristics play an important role in the link between demographic and economic sub-systems, though the degree of reliance on variables such as these varies from one Bachue model to another.

The logic of the endogenisation of the demographic variables partly rests on the existence of feedbacks from these variables to the behaviour of the economic-demographic system as a whole. But it also depends on the hypothesis, established or assumed, that each of these variables is to some extent responsive to socioeconomic change. These are empirical issues, and consequently the endogeneity of some of the demographic variables varies from one country model to another.

II. POPULATION ACCOUNTING AND INTERTEMPORAL STRUCTURE

The core of the demographic sub-system consists in tracking the population over time. The primary disaggregation of population is by age and by sex, criteria which are present in most behavioural relationships, either explicitly or implicitly. Certain age classes are more likely to migrate, and migration can also be sex-specific; schooling affects particular age groups; over-all fertility depends on the age structure of the female population; young entrants to the labour market are less likely to find a job easily; female school enrolment is less than for males, and so on.

The tracking of disaggregated population cohorts raises problems of data management, but it is in principle the simplest approach to obtaining population cohorts of a given age and sex at different points in time. Although techniques exist to describe population's age and sex structure through distribution functions, the approximations which such functions involve are risky to introduce in a long-term demo-economic system undergoing substantial changes. The same reasoning has also led to a breakdown of population both spatially and by education level—these two disaggregations corresponding to structural features of the economic sub-system—with the sub-populations so identified tracked over time.

The population accounting consists in tracking over time identified population cohorts broken down by age, sex, location, education level, and whether or not enrolled in school. It is essentially an elaborate cohort component projection method, establishing that population at time \( t \) is equal to population at time \( t - x \) plus births minus deaths plus net in-migration occurring during the period \( x \), together with an adjustment which allows for changes in the other dimensions of disaggregation (school enrolments, drop-outs, etc.). Thus

\[
P(t) = A(t-x) \cdot P(t-x)
\]

The elements of matrix \( A \) are the subject of this paper. From the accounting point of view, they are fertility, survival, migration, enrolment, drop-out, labour force participation, nuptiality and headship rates, each broken down according to the basic disaggregation criteria: age, sex, location, education level.

With respect to the time period of the models, early work on a prototype Bachue used a "systems dynamics" approach [6]. The underlying structure of the model in this approach is a set of differential equations; the period of time in which such equations are solved is short relative to the simulation period as a whole, and so the changes which occur in model variables in one time period are small relative to total changes in a given simulation. This first model used as basic time period an exogenously modifiable fraction of one year. In practice, however, it and subsequent models have tended to use one year as the basic time period, which appears to be short enough to describe most phenomena considered in a long-term oriented model, in a recursive way, and long enough to avoid certain types of other short-term (e.g., seasonal) relationships, and unnecessary computation.

The fact that the basic time-period of the model (e.g. one year) does not correspond to the period during which a variable is measured or to which it refers (e.g., population broken down by five-year age groups), does not matter in the "systems dynamics" approach but leads to the introduction of unfamiliar notions. In Bachue-
Philippines, the fraction of five-year age groups transferring to the next one during one year is computed through a "maturation rate". More examples were found in Bachue 1 [6].

III. FERTILITY

(i) Measurement of Fertility

The measure of fertility used as a dependent variable in the various Bachue models differs considerably between applications. Given that the population is broken down by age and sex, the measure of fertility which is most consistent with this structure is age-specific fertility rates. Such rates already allow through their disaggregation for one of the factors, i.e. changes in age structure, which are frequently recorded as responsible for changing over-all fertility levels. If age-specific marital fertility rates are used, changes in age at marriage are also taken into account.

However, more aggregated measures have been used as the dependent variable in certain modelling applications. For instance, the Philippine model uses the gross reproduction rate, while the Brazilian model uses three age-specific fertility rates. Clearly some information is lost in these approximations (mortality pattern in the GRR, nuptiality pattern if the index is computed on the whole female population).

An interesting alternative procedure is to treat the parameters of a fertility distribution function as dependent variables [13]. This approach, which is analogous to the derivation of age-specific mortality rates from life expectancy at birth using life tables, has been applied in Bachue-Yugoslavia.

(ii) Explanatory Variables

Endogenisation of fertility in an economic-demographic model of the Bachue type is largely based on a theory of household behaviour. Fertility is regarded as resulting from a conscious decision, which takes into account the quantitative advantages and disadvantages of child-bearing. Such an approach can be seen in the utility-maximisation framework of the "new home economics", though this is not necessary, since most of the implications of the former can be derived from simpler analyses. Socio-economic characteristics of the household are seen in this framework as major determinants of its fertility behaviour. Important variables include the education level of the mother and father, the mother's past and present economic activity, household income, child mortality, and the education or desired education of children. The gains from child-bearing also depend on the foreseen economic activities of children, and the extent to which parents benefit from these activities. A considerable literature exists on the precise specification of models of this type.

The explanation of fertility behaviour is therefore essentially microeconomic. Such an approach does not normally capture variables such as norms, tastes, psycho-
but mostly because of the way data are usually collected. Household surveys rarely gather, at the same time, information at a higher level of aggregation, and it is not unfortunate. Cross-sectional variations in fertility behaviour might be small, relative to changes and for large deviations from this norm to have a considerable random component.

In practice, therefore, the micro relationships are necessarily treated for modelling purposes, as estimates of the corresponding macro relationships, which means that the omission of some, at least, of these more aggregated variables is particularly unfortunate.

The method of estimation of micro relationships affects the sensitivity of the functions concerned. For instance, when household surveys are used, national or cross-sectional variations in fertility behaviour might be small, relative to changes over time, because of the tendency for fertility patterns to cluster around the norm, and for large deviations from this norm to have a considerable random component. As a result only small variations in fertility are explained in a cross-section function. By contrast, when a fertility function is estimated on the basis of aggregated data, pooled cross-sections or international cross-section data, clear distinctions between relatively homogeneous groups will tend to be associated with fertility differentials, and explanatory variables will be more of a structural nature. This will tend to be reflected in a more sensitive fertility function. The choice of specification therefore finally relies as much on hunch and intuition as it does on econometric refinement.

IV. MORTALITY

(i) Measure of Mortality

The representation of mortality in the Bachue applications uses the normal procedure of demographic projections: a synthetic index such as life expectancy at birth, or the probability of a birth surviving to age three, is projected and used to recompute the components of the index, i.e. the age-specific survival rates. The accuracy of the mapping of the overall index on to age-specific rates varies. In the Brazil model, national life tables were constructed whereas in the Philippines and Kenyan models standard Coale-Demeny life tables were used. In Yugoslavia the relationship between mortality patterns below and above 5 years of age was found to

logical factors, values, etc; nor are structural socio-economic relationships, such as patterns of class behaviour, allowed for. The reason for not considering such elements in fertility behaviour is not so much that microeconomic theory and other theories have not been linked together, or even that they conflict with each other, but mostly because of the way data are usually collected. Household surveys rarely gather, at the same time, information at a higher level of aggregation, and it is not always easy to link aggregate data from other sources with household survey data. In practice, therefore, the micro relationships are necessarily treated for modelling purposes, as estimates of the corresponding macro relationships, which means that the omission of some, at least, of these more aggregated variables is particularly unfortunate.

The method of estimation of micro relationships affects the sensitivity of the functions concerned. For instance, when household surveys are used, national or cross-sectional variations in fertility behaviour might be small, relative to changes over time, because of the tendency for fertility patterns to cluster around the norm, and for large deviations from this norm to have a considerable random component. As a result only small variations in fertility are explained in a cross-section function. By contrast, when a fertility function is estimated on the basis of aggregated data, pooled cross-sections or international cross-section data, clear distinctions between relatively homogeneous groups will tend to be associated with fertility differentials, and explanatory variables will be more of a structural nature. This will tend to be reflected in a more sensitive fertility function. The choice of specification therefore finally relies as much on hunch and intuition as it does on econometric refinement.

(i) Measure of Mortality

The representation of mortality in the Bachue applications uses the normal procedure of demographic projections: a synthetic index such as life expectancy at birth, or the probability of a birth surviving to age three, is projected and used to recompute the components of the index, i.e. the age-specific survival rates. The accuracy of the mapping of the overall index on to age-specific rates varies. In the Brazil model, national life tables were constructed whereas in the Philippines and Kenyan models standard Coale-Demeny life tables were used. In Yugoslavia the relationship between mortality patterns below and above 5 years of age was found to
be poorly represented in model life tables. Life expectancy at age five did not vary much between regions, and future mortality declines were unlikely to be large, this was not the case for ages 0 – 5. Therefore, these two age groups have been treated independently and each used to locate its corresponding segment in the model life tables. Moreover only child mortality has been endogenised.

Sex differences are allowed for in this approach in so far as mortality rates differ between males and females, and also in so far as life expectancy at a particular point in time is differentiated by sex. It has in general been assumed that the initial male/female differences in life expectancy will remain constant over time or that they are adequately represented by the patterns in model life tables. These are rather crude assumptions, particularly in countries where female life expectancy is less than male life expectancy, a difference which no doubt reflects socio-economic rather than biological factors and which is therefore likely to change.

(ii) Explanatory Variables

With the exception of the Kenyan model, where a fuller endogenisation was attempted, mortality has only been explained by aggregate socio-economic variables without taking into account the factors directly affecting the health of individuals. The main reason for this is that the latter effects are difficult to assess, except where major endemic diseases exist. Rather than identifying the immediate causes of survival, therefore, the explanation of mortality has been couched in terms of socio-economic correlates. Low levels of mortality are found in higher income countries, and are associated with extensive medical infrastructure, with urbanisation and with higher educational levels. The distribution of the population with respect to these variables, and the degree of non-linearity in the relationships concerned, are also important.

This approach raises a serious issue in policy modelling; viz. since the direct effects of health programmes on health are difficult to observe, their impact in the model has to be allowed for through exogenous changes in the dependent variable rather than through the manipulation of independent policy instruments. For example, it would not be possible to assess the impact in the model of water supply programmes or the provision of intermediate level medical services, unless the impact of the programmes on health is already quantified. The problem is similar to that raised by the exogenous way in which the impact of family planning on fertility is represented in all Bachue models with the exception of the application to Yugoslavia. This lack of endogeneity of policy measures is unfortunate because, while long-term trends in mortality are clearly associated with long-term socio-economic changes, specific programmes certainly have an impact in the short term.

In Kenya, a comprehensive household survey was used to identify the determinants of sex-specific survival probability up to age three. Variables of direct interest to health programmes, such as access to piped water, the availability of health infrastructure, and the prevalence of endemic disease, were included in addi-
tation to socio-economic characteristics of the household, such as education level, income, and housing conditions. However, a simplified equation has been retained in the model with an exogenous trend accounting for all policy-related variables.

V. MIGRATION

(i) Dependent Variables

There are various migration flows which can be identified and which are represented in the Bachue models. A first disaggregation is between external (international) migration and internal migration. Among the latter flows, one may distinguish intra- and inter-regional migration, the nature of which will of course vary with different definitions of regions. Mobility between economic sectors may also involve physical movement of individuals, and this type of economically-based disaggregation will lead to a classification of migration which differs from that based on regions.

International migration is included in the Bachue models only where major flows can be identified. In the Philippines, for instance, emigration of educated urban Filipinos is allowed for. In Yugoslavia, both emigration and return migration (and the associated financial flows) are incorporated in the model.

Internal migration between regions is treated in much greater depth in all models. Where the regions are rural-urban, this type of migration covers both geographical movement and, usually, sectoral labour mobility, in particular out of agriculture. In the Yugoslav case, where the rural region is identified with private agriculture, migration necessarily implies sector mobility. In other models, rural out-migrants are usually, but not uniformly, agricultural in origin. In addition to a rural-urban disaggregation, the Yugoslav model is divided into 8 administrative regions. Current development of the Brazilian model also includes its disaggregation into 5 geographically specified sub-models. In the Yugoslav application there is also a significant contribution of the rural population to urban production, through a commuting phenomenon, which then constitutes another type of spatial migration. In none of these models, however, are seasonal and short-term migrations explicitly represented in the model.

Intra-regional movements are considered in an economic rather than a spatial sense. For instance, sectoral mobility between different labour categories is possible, and this type of mobility is an important characteristic of the development process. However, the geographical distribution of population is not modelled to the same level of detail — for instance, the distribution by size of cities is not considered.

Internal migration flows can be treated either as gross or as net. There is however a significant advantage from treating them as gross as the determinants of migration in different directions are not necessarily symmetrical. This is particularly true of rural-urban migration as compared with urban-rural.

(ii) Independent Variables

While behavioural factors, of course, enter into the determination of international migration, there is much greater scope for physical controls than is the case with internal migration. It is therefore reasonable to treat international migration as exogenous. Internal migration, on the other hand, is seen in all the Bachue models as an adjustment mechanism redistributing a factor of production according to the relative advantages of regions of origin and destination, though the outcome is by no means equilibrium, or even necessarily a movement towards one. Certain groups of the population are more sensitive to changes in relative socio-economic conditions than others, and so there needs to be a careful disaggregation of the migration relationship. But in all models the basic explanation is in terms of a balance between push and pull factors.

Migration is treated as a decision by an individual or a household, resulting from the comparison of present and possible future economic conditions in the place of current residence, and those in other regions. The income differential between regions is usually the main explanatory variable. A number of different variables are introduced to represent expectations: the unemployment rate and the inequality of income distribution are two of these. There are different classes of migrants, in part corresponding to particular segments of the labour market. Unskilled migrants, for instance, may be concerned exclusively with the possibility of finding a job in traditional/informal sectors and with the income they can expect in such a job; women will take account of the availability of “female” jobs; and so on. The propensity to migrate may also differ according to sex, age, education, marital status, and family situation.

While this approach to migration is common to all the models in the Bachue family, the nature of the migration process is country-specific. Even if the explanatory variables are similar in different countries, their measurement, their conceptualisation and their quantitative evaluation, are likely to vary considerably. The main differences in specification from model to model concern the way individual characteristics and environmental factors have been treated. In some cases these have been built into the same function (Brazil, Yugoslavia); in others (Kenya, Philippines) two separate functions have been estimated, one explaining the determinants of aggregate migratory flows, and the other determining the probability that an individual will migrate, given the overall flow.

Data on migration, detailed enough to allow functions of this latter type to be estimated, are only available from household surveys. These provide both time series (recall) data, which may be more or less complete and which provide some information on the economic condition before migration, and cross-sectional information. Even so, it is often difficult to consider historical and structural factors. In addition, temporary factors with a large but unascertained influence on migration are difficult to isolate out. Among problems encountered, notice can be taken of the
### Table 4

**Migration**

<table>
<thead>
<tr>
<th></th>
<th>Brazil</th>
<th>Kenya</th>
<th>Philippines</th>
<th>Yugoslavia</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>External</strong></td>
<td>None</td>
<td>None</td>
<td>Policy variable for some age-education groups.</td>
<td>Permanent out-migration rates by age, sex, education and regions: policy variable. Temporary outmigration and return migration rates of workers and dependents by age, sex, education, region, A/NA: policy variable.</td>
</tr>
<tr>
<td><strong>Internal Dependent Variables</strong></td>
<td>Urban household head's probability to come from rural areas (no sex distinction).</td>
<td>Overall net R/U migration rate. Age, sex, education specific net migration probabilities derived from overall migration rate using fixed relative rates.</td>
<td>Gross migration probability (macro) by sex between R/U &amp; gross migration propensity (macro) by sex between disaggregate R/U areas. Gross migration rate combines both dep. variables multiplicatively with population weighting factor. Male and female rates assumed to be equal in simulation.</td>
<td>Gross migration rates from region i to A/NA sector of region j, by age, sex, education. Net migration rates out of (A) by age, sex, region. Inexistence of certain types of flows between and within regions of A/NA populations, assumed because of over-determination.</td>
</tr>
<tr>
<td><strong>Explanatory Variables</strong></td>
<td>Income differentials between regions, education, number of children, age.</td>
<td>Micro: education level Househoold income per AE differentials; R/U modern sector wage rate &amp; employment probability.</td>
<td>Micro: age, education level, marital status Macro: Income differentials, income distribution, R/U.</td>
<td>For inter-regional flow to NA sector: average wage, job creation rate, unemployment &amp; female LFPR differentials, % women married in i, regional dummies. For inter-regional flow to (A) sector: land size, differentials % women married in i, regional dummies. For net out-migration: NA average wage, job creation rates, unemployment and female LFPR % women married and land size in i, regional dummies.</td>
</tr>
<tr>
<td><strong>Comments</strong></td>
<td>Logit estimation. Household survey and population census data. Adjustment of overall migration rate to census figure.</td>
<td>Bi-log overall rate function based on district level data. Non-linear functions for relative rates derived from national household survey. Return migration to rural areas allowed.</td>
<td>Linear micro function, log-linear macro function. Data: national household survey; micro based on individual data; macro on aggregated data at province level. No adjustment of computed values.</td>
<td>Linear specifications. Most explanatory variables are education and often age-sex specific. Child rates (below 15) function of childbearing age female rates. Population census and socioeconomic aggregate data for 79 demographic regions. Adjustment of intercept &amp; dummies to region and A/NA observations.</td>
</tr>
</tbody>
</table>

**Note:** See footnote to Table 1.
insufficient description of the household's socio-economic environment (e.g. land structure) and the fact that certain macro-level factors which are expected to affect migration are measured after migration whereas migration rates can influence them (e.g. employment structure, unemployment rate). Moreover, many public policies, such as land reform, introduction of new crops in certain regions, rural public work programmes, or physical controls (often of temporary or variable nature), can hardly be assessed without longitudinal surveys. There are thus some factors which limit the validity of the explanation of migration followed in the Bachue models. In particular the inadequate representation of structural factors suggests a direction for improvement.

VI. Nuptiality

Nuptiality is an important variable of the demographic system in at least three ways. Firstly, it has close links with fertility: to the extent that all births take place within marriage, it determines, in combination with marital fertility, the level and age pattern of over-all fertility. Secondly, the economic activity of married women tends to differ considerably from the activities of unmarried women. Thirdly, marriage is a component of household formation. However, the treatment of nuptiality in the Bachue models varies much more than is the case for fertility, mortality and migration. Actually, nuptiality is only explicitly considered in Bachue-Philippines and Bachue-Yugoslavia, where legal marriage is traditionally the only form of union which leads to child bearing.

(i) Dependent Variables

The dependent variables can either be the proportion of married women in each age group, or, more commonly, the parameters of a function which determines the distribution of married women by age group, a procedure followed in both models which consider nuptiality. There are relatively few comparative studies of the latter distribution, and those which have been carried out have tended to use a distribution function developed by Coale [13].

(ii) Explanatory Variables

In a “new home economics” model, the decision to marry would be regarded as a rational decision based mainly on economic factors. In particular the education of the woman is likely to delay marriage, because schooling raises the opportunity cost of marriage while at the same time reducing its benefits because more educated women are potentially economically more independent. Similar patterns of relationships can be observed between economic activity and marriage patterns in that the work of girls and young women raises the returns to the family of origin, which therefore has an interest in delaying marriage. It is difficult to translate this model directly into the explanation of parameters of the distribution function, but approximations can be made.
Of course, only economists could think that the decision to marry is basically an economic one. Norms, tastes and culture also intervene; indeed, are arguably dominant. One consequence is to make international comparative analysis of the age at marriage almost meaningless, because the content and functions of marriage vary so much between countries. Unfortunately, most of these non-economic factors are difficult to analyse quantitatively, so that in practice determinants of marriage rates tend to be based exclusively on the more readily measurable variables such as education and labour force activity.

VII. HOUSEHOLDS

Under this heading we include three issues: determining the number of households, their demographic characteristics in terms of size and age distribution of their members, and their socio-economic characteristics. There are conceptual problems in treating households as integrated units with definite boundaries, but we look to others for advances with respect to conceptualisation [29] and set the problem on one side here.

Little quantitative evidence exists on these various relationships. The little empirical work which has been done, mostly using population censuses, indicates that there are sizeable variations in headship rates between different levels of development, but also between cultures and between countries with different demographic characteristics. Only one model endogenises headship rates, that for Kenya, where the rates (specific to women aged 20–59) are made a function of the adult sex ratio. In other models group-specific headship rates are kept constant.

Demographic Composition of Households

A number of aspects of economic behaviour depend on the size and age structure of households, including household consumption patterns, the participation of individuals in economic activities, and fertility (influenced by the parity level, sex ratio of children and the presence or absence in the household of people able to take care of children). In many cases, household income is also relevant, but these examples show the need to include demographic variables as well. In fact we would argue that the treatment of households is one of the most important issues in analysing economic-demographic interactions in that it acts as a link between a number of different sub-systems. Nevertheless, little work has been done in this field, and the explanation of household formation and demographic characteristics is therefore relatively crude in the Bachue models.

The treatment of household composition in the Bachue-Philippines and Bachue-Kenya models is the most elaborate among the models considered here. Both models start with an initial distribution of households by income group and demographic composition (size and age structure). In the former application, the implied headship rates are maintained constant over time while changes in the size

| Demographic Composition of Households |

<table>
<thead>
<tr>
<th>Size and Age Structure</th>
<th>Income Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children</td>
<td>Adults</td>
</tr>
<tr>
<td>Presence</td>
<td>Absence</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variables</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income</td>
<td>Household income, household size, household age structure</td>
</tr>
<tr>
<td>Size</td>
<td>Household size distribution by income group</td>
</tr>
<tr>
<td>Age</td>
<td>Household age structure distribution by income group</td>
</tr>
<tr>
<td>Occupation</td>
<td>Household occupation distribution by income group</td>
</tr>
</tbody>
</table>

Table 6

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Philippines</td>
<td>Household composition and socio-economic characteristics</td>
</tr>
<tr>
<td>Yugoslavia</td>
<td>Household composition and socio-economic characteristics</td>
</tr>
</tbody>
</table>

Dependent Variables

- Number of females 20 and above in household
- Average household size
- Average income of household in total
- Headship rates by age and sex

Explanatory Variables

- Household size distribution of total labour force
- Mean household income
- Household size function of income
- Analysis of household formation and changes in distribution

Comments

- See footnote to Table 1.
and composition of households by income group are assumed to affect all types of household similarly (as noted below, different sub-groups are identified). This is in contradiction with the micro theory of fertility behaviour, in which the nature of the household will influence the fertility pattern. Changes in socio-economic conditions should therefore act on fertility in a way which is differentiated by household type, so that differential changes in age-sex structure will occur. Strict proportionality of changes is therefore likely to be a rather poor assumption and work to improve this approach would be desirable, for instance developing further the link introduced in Bachue-Kenya, between the size distribution and income distribution of households.

Other Classifications

Income is not the only criterion for the classification of households. Another important classification in certain applications of the Bachue models is the distribution of households according to the employment status of the head. Here again the initial conditions are easily determined, but the evolution over time can be complex. In Bachue-Philippines and Bachue-Kenya, where these issues are treated in more detail than in other models, the occupational distribution of household heads is known at the origin, and is assumed to change proportionally with the over-all occupational distribution of employment. Thus a decrease in over-all agricultural employment implies a decrease in the number of agricultural households.

Associated with the occupational distribution of household heads, the distribution of non-heads according to the head’s occupation is also required for household income distribution purposes in cases where household members other than the head frequently have a remunerated economic activity.

VIII. LABOUR FORCE PARTICIPATION

(i) Dependent Variable

The measurement of the labour force for modelling purposes raises a host of conceptual problems, discussed in some detail in Standing [44]. The use of inappropriate labour force concepts not only means that the measurement of labour supply is inexact, but also can seriously affect the behaviour of the labour market sub-system. For instance, neglecting to differentiate labour demand between male and female workers, and between different age groups (as is indeed the case in all the Bachue applications except for the forthcoming regionalised application to Brazil) is likely to produce some misleading results. Care is therefore required and there may be a case for incorporating several alternative definitions of labour supply (although this has not been done in any model so far).

For modelling purposes, it often appears that certain categories — notably, adult males — have labour force participation rates which are very similar across.

(ii) Determinants

The determinants of labour force participation include elements of household economics, demographic characteristics of the individuals and labour market conditions. A typical equation would explain labour force participation, disaggregated by sex and age, as a function of the education level, marital status, the number of children or the age of the youngest child, position in the household, household head’s occupation, industrial structure, minimum wage for a given education or skill level, household income, unemployment and the rate of job creation. See Standing [44], and Standing and Sheehan [45].

In the Bachue models, cross-sectional household survey data have usually been used to estimate such equations, sometimes aggregated at the regional level. The use of such data may restrict the inclusion of market-related variables such as industrial structure, previous labour absorption, or the unemployment and wage rates, because there is likely to be insufficient variation in the sample or because these variables have just not been measured in the survey. In consequence, many studies of labour force participation in developing countries omit demand factors. In the Brazilian model, this omission was considered sufficiently serious that labour force participation rates are determined by long-term trends rather than by a participation function. At the other extreme, in both the Yugoslav and Kenyan applications, market variables have received much attention in specifying participation functions.

One consequence of the absence of demand factors in some of the models in explaining labour force participation rates is that they tend to be too insensitive to market changes, causing instabilities elsewhere. On the whole, instability due to specification error is not a major problem in large and highly endogenous economic-demographic modelling. Behavioural equations in such models generate a series of feedback mechanisms which allow the system to perpetuate itself or to smoothly change in the direction of market or social pressures. In the case of labour force participation equations, however, changes on the demand side have an impact primarily on work and incomes rather than on the size of the labour force and unemployment, making wages relatively unstable. This can be partly overcome if an income variable is incorporated as a determinant of labour force participation, though it is not always obvious to what extent the cross-section relationship estimated will adequately reflect changes as incomes rise over time. There are indeed reasons to believe that income, as a determinant of labour force participation, should be expressed in relative rather than absolute terms.
IX. EDUCATION AND SKILL

(i) Dependent Variable

Education is one of the basic criteria used to disaggregate the population. Education levels are important direct explanatory variables of all demographic events, and either education or skill is also an important factor in labour market behaviour. By education is usually meant the grade completed or the level obtained, irrespective of the number of years spent in school. However, in all applications of the Bachue models, except to Kenya, the number of years of schooling and the grade obtained are equivalent, as no repeating of grades is allowed for.

(ii) Determinants of Education Levels

The education level of the population is usually determined by tracking school-age children through the education system in a manner similar to the tracking of population cohorts through their life cycle. The dynamic behaviour of the system is therefore determined by transition probabilities from level to level, which in turn depend on enrolment, completion, drop-out and (in Bachue-Kenya) repetition rates. Such rates have been shown to be a function of the education facilities available, certain specific government policies with respect to education, and characteristics of the child and the child’s family which determine behaviour with respect to investing in human resources development. In most of the current applications of the Bachue models the supply side (i.e. the education facilities and government policy) is treated as the main determinant of such rates. In most developing countries, this approach is probably acceptable as a first approximation. Existing household survey data were inadequate for the testing of more elaborate models which incorporate both supply and demand factors. With some differences between the models, education rates are thus considered policy variables or determined by over-all government expenditure patterns.

X. CONCLUSIONS

Drawing general conclusions on the treatment of demographic variables in the Bachue models with a view to simplifying the construction of similar models, under other circumstances, is difficult. The main reason, in our opinion, is that a systemic model, i.e. the representation of the structure of a system, is exclusively defined spatially and temporally, hindering the transfer of parts of one demo-economic system to another and even the comparison between them. For instance, it would be foolish to compare the structure of the demographic system of Bachue-Yugoslavia with that of Bachue-Kenya with a view to determining the better one. Indeed, the validity of the representation of a system mainly depends on its ability to simulate the dynamics of measurable phenomena, such as the population growth rate or the quantity and quality of the labour force, and not so much on the underlying theories.
<table>
<thead>
<tr>
<th>Dependent Variables</th>
<th>Brazil</th>
<th>Kenya</th>
<th>Philippines</th>
<th>Yugoslavia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability to belong to one of the four education levels, R/U.</td>
<td>Enrolment, repetition and completion rates by age, sex, R/U.</td>
<td>Conditional rate of success in primary and secondary determining three education levels, R/U.</td>
<td>Enrolment rates in higher primary, two secondary, tertiary by sex, A/NA</td>
<td>Enrolment rates in lower primary, drop-out rates from all levels by sex, region, A/NA.</td>
</tr>
</tbody>
</table>

| Explanatory Variables | LFPR, household income public expenditures for education, sex, age. | None | None | Enrolment in lower primary and drop-out rates exogenous and/or policy variables. For higher primary rates: young age dependency, NA father’s education, A child LFPR, regional dummies. For secondary: labour supply, education level, employment growth, unemployment, child and female LFPR, % young women married, father’s education, regional dummies. For tertiary: % general secondary graduates, % tertiary level labour supply, father’s education, secondary education, unemployment. |


Note: See footnote to Table 1.
Another factor which does not facilitate comparisons between models is that each country application has been developed under specific conditions determining particular features of the model. Data availability is one of these conditions. The presumed interest of potential users, notably the interest and involvement of policy makers in the outcomes of the model, is another condition shaping each model.

Given these problems of comparability, the conclusions will be limited to some general comments on endogeneity, on the demographic accounting and on certain issues raised by the estimation of behavioural functions and their introduction in the model.

(i) Endogeneity

While the endogenous nature of certain demographic elements of a demographic model is clear, the construction of the Bachue models has shown that there are no precise rules valid for all cases, i.e. adequate to describe different socio-economic systems and to satisfy all possible users. As mentioned above, there is considerable variety in the way characteristics of the population have been represented in the various applications. To some extent at least, this variety reflects the mechanisms of the system described, and the validity of the representation of the structure should be considered the main criterion to decide on the endogeneity of an element.

In Bachue-Brazil, marital status has not been introduced in the present version of the model. Actually at least three types of marital status would have to be considered: single, legally married and consensual union. The other sub-systems of the model do not take into account the differences between these three marital statuses, so that this distinction would have no consequences and could therefore be redundant. But a more detailed analysis of the generation of poverty might be needed to remedy this omission. It has been shown elsewhere that in their precarious situation, women living in consensual union are more likely to be forced to work than those in regular marriage, and when working, often paid less than the minimum wage. There is therefore a case for disaggregating labour market and income distribution systems by marital status. This is a case where the level of disaggregation of one sub-system of the model has direct implications for the level of disaggregation of other sub-systems. Does this suggest that marital status should be an endogenous element, or is it sufficient to consider it a parameter? It depends partly on the structure of the system, partly on policy objectives in model design and partly on the availability of data. The first two cases are rather straightforward. In the third, if an element is known to be endogenous but data are not available to measure quantitatively its dependence on other elements of the system, the question for model builders is whether to retain the element as a parameter or to use simulation techniques to represent the assumed relationship. Although simulation has a poor image in many circles, the authors of the present paper and probably most of those who have worked on Bachue models would favour such an approach. It is no doubt risky to work with best "guesstimated" relations, but assuming unchanged coefficients is even less reliable. In the marital status example above, an acceleration of migratory flows might increase the number of women living in consensual union, and thereby have a considerable secondary impact on employment and poverty. Keeping a constant ratio between the types of marital status would hide this phenomenon.

(ii) Accounting

The Bachue models have often been criticised for the size and complexity of their demographic systems, either in comparison with those of the economic system, or in comparison with the magnitude of changes occurring in the short and medium term. For medium-term models, it is often argued that a full demographic accounting and behavioural system are not necessary and could be replaced by simple extrapolation of the main elements. This point is valid to the extent that the model is essentially aimed at broad relationships between the economic and the demographic systems, but there are several practical reasons in favour of a complete system.

Firstly, in contrast with economic models, a complete demographic accounting is probably easier to develop than an abridged version. If the model is simplified too much, certain elements, especially those which can have short-term effects and are therefore of greater relevance to policy makers, tend not to respond to changes in the system. For instance, this might concern the proper accounting of new entrants in the education system, the sex ratio of entrants into the labour force or the tracking of spatially disaggregated populations.

Secondly, the outcomes of a detailed accounting, while perhaps not relevant to the medium-term dynamics, may be much more relevant for partial planning. It is of interest to education or health planners to know, rather accurately, the eligible population according to certain characteristics.

Thirdly, a complete accounting can be adapted rapidly to various needs by dropping, ex post, redundant information. In contrast, if additional information has to be generated by extending a simplified system, a long and tedious programming task is often implied.

Meanwhile, there are obviously valid arguments (cost, ease of interpretation, etc.) in favour of small systems but the experience with model construction tends to show that the redundancies can only be eliminated after and not before the sensitivity of the model has been fully tested.

(iii) Estimation and Use of Behavioural Equations

The estimation of behavioural relations raises two main questions: the source of data and the type of function used in the programme. The Bachue models have used a mixture of international and national cross-section data sources to estimate the behavioural functions of the demographic sub-system. In so far as national cross-section data sources are used, either individual or aggregated data can be used. Time series or a combination of cross-section and time series can be used and in
dummy variables which would account for the differences between the sub-groups. The use of international cross-section has almost always resulted from the lack of national data sources. The main exception is the fertility function of Bachue-Philippines where international cross-section data were preferred in view of the inadequate dynamics of an equation estimated on national data. Meanwhile international cross-section data suffer from many drawbacks. Firstly, the range of explanatory variables is very limited, especially in comparison with the number of elements available in the model itself. A complete “explanation” of any demographic phenomenon is out of the question. Secondly, international cross-section data give an average measure of phenomena, whereas at the national level regional and inter-group differences may be crucial for an adequate analysis. Thirdly, international cross-section data, because of their very nature, are highly multicollinear. This makes it difficult to separate out the effects of different independent variables, and also the coefficients of explanatory variables may actually be reflecting the influence of other collinear variables not in the model. If the model is policy-oriented, designed to look at the effect of shocks on the structure, explanatory variables are not likely to move parallel any more. In this case, each coefficient should represent the sole effect of the variable concerned so as not to bias the policy-simulation results.

Aggregated national data can present the same drawbacks as international cross-section data, though the bias is smaller and depends on the criteria of aggregation: for instance, in structuring data by income groups, one often introduces a systematic bias. Micro data (i.e. data at the level of the decision unit, whether individual or household), do not present these drawbacks but because of the magnitude of the error term, the level of statistical significance of the equation estimated can be very low, which casts doubts on its structural validity.

For purpose of sensitivity analysis of model behaviour, and also to make model structure more coherent, it would be desirable to have a smooth pattern of coefficients, e.g. by deriving from the original coefficients a consistent pattern over age groups.

Related to this issue, there is the problem raised by the estimation of behavioural functions with a view to using them in a much wider context. Should one estimate the best possible function and use it as such in the simulation or is it better to lose some capacity to explain the variance in order to control the simplicity of the explanation? This problem is illustrated by the estimation of equations in Bachue-Yugoslavia where the basic disaggregation is by republic and mode of production (private and social). In this case, there is the possibility (a) to fit one unique specification of the function to the whole sample and to compute the value of the dependent variable in each sub-group, from the values of the explanatory variables in that sub-group; (b) to introduce in the specification of the general model dummy variables which would account for the differences between the sub-groups (this is different from the previous approach as differences between sub-groups are predetermined to the model); and (c) to fit the best specification to each case, perhaps reaching rather large differences from one specification to another. The first approach (unique general model) can be dangerous as the necessary calibration of the equation on the initial data can hide very large specification error. The second approach at least draws attention to the differences though it does not explain them. Indeed, if dummy variables on the intercept and on the slope of the explanatory variables are significant, they actually compensate for specification errors, such as the omission of institutional constraints or other macro effects. The third approach is obviously preferable for understanding a demographic phenomenon, but does not necessarily provide the best behavioural equations for a demographic model; often such equations are too complex, containing numerous exogenous variables for which a satisfactory evolution over time has to be decided upon. Some of these difficulties derive from data inadequacies. However, the lack of data is probably no longer the major constraint on the construction of such models (although data are sometimes deficient in quality and in comprehensiveness, and often biased in the way they have been collected to test one *a priori* hypothesis). Nor is the extent of economic and demographic theory a major limitation — indeed as far as theory is concerned the problem is the “trop-plein” rather than the “vide". The major problem in building demo-economic models lies more in the formalisation (including the quantification) of all sorts of partial theories and empirical knowledge into a coherent and global frame, rather sophisticated in some ways, but also highly imperfect; and after this formalisation, in converting this conceptual structure into a tool which can be effectively utilised for understanding a social system, and for evaluating attempts to modify that system through policy interventions.

REFERENCES


