MEMORANDUM
RM-5405-RC/AID
DECEMBER 1967

A FAMILY PLANNING HYPOTHESIS: SOME EMPIRICAL EVIDENCE FROM PUERTO RICO

T. Paul Schultz

PREPARED FOR:
THE RAND CORPORATION AND THE
AGENCY FOR INTERNATIONAL DEVELOPMENT

The RAND Corporation
SANTA MONICA • CALIFORNIA

MEMORANDUM
RM-5405-RC/AID
DECEMBER 1967

A FAMILY PLANNING HYPOTHESIS: SOME EMPIRICAL EVIDENCE FROM PUERTO RICO

T. Paul Schultz

This research is supported by the Agency for International Development, Department of State, under Contract No. F44620-67-C-0050 and The RAND Corporation. RAND Memoranda are subject to critical review procedures at the research department and corporate levels. Views and conclusions expressed herein are nevertheless the primary responsibility of the author, and should not be interpreted as representing the official opinion or policy of the Agency for International Development or of The RAND Corporation.

DISTRIBUTION STATEMENT
Distribution of this document is unlimited.

The RAND Corporation
1700 MAIN ST • SANTA MONICA • CALIFORNIA • 90406

Published by The RAND Corporation

PREFACE

Convenient, cheap, and universally available birth control devices are often regarded as a total solution to the population problem. Actually, this may be a necessary part of the solution, but only part. What if parents continue to want more children? This neglected possibility is explored here. Can differences in birth rates be interpreted partly as reflecting differences in the number of births parents want, and if so what aspects of the parents' environment are responsible for their different preferences?

As a tentative step toward applying the tools of economic analysis to population growth, activity, and migration, this study investigates the determinants of birth rates. It is assumed that people tend to adjust their behavior for their own betterment when their environment affords them new opportunities and imposes on them new limitations. This tendency toward behavioral adjustment may be neither prompt nor complete, because humans are bound by inertia and guided by imperfect information and because change is incessant. The usefulness of this working hypothesis is ultimately judged in terms of its ability to predict behavior. Having thus adopted the methodology of economics, we find it difficult to compare our findings with those in parallel fields where disciplinary approaches differ radically. Mainly for this reason, no survey of the extensive literatures of demography, sociology, and anthropology is made here.

Other related studies are underway at RAND dealing with the simultaneous determination of reproductive behavior, economic activity, household income, and regional migration. This initial Memorandum does not attempt to specify or estimate these important areas of interaction. Also, a future Memorandum will discuss the implications of these findings for public policy, and will seek to trace the consequences of alternative development strategies to the rate of population growth, its regional distribution, the use of and investment in human resources, and the functional and personal distribution of income.

The author acknowledges financial support for this study from the Latin American research program sponsored by The RAND Corporation's own funds and directed by Herbert Goldhamer, and from the Agency for International Development. This study represents a pilot attempt to develop a methodology. It is planned that this framework will be applied to Colombia in the context of the RAND Colombia study, financed largely by AID. Richard Nelson has contributed most to this research by his encouragement and acute criticisms, and other RAND colleagues and consultants, Herbert Goldhamer, John McCall, Edward Mitchell, Marc Nerlove, R. E. Park, and James Schlesinger have helped to improve the style and substance of this Memorandum. Any shortcomings that persist are the author's.

SUMMARY

This Memorandum sets forth the hypothesis that the frequency of births in a population can be understood in terms of three groups of factors that influence parents' desires for births. First there is a family size goal or a number of surviving children that parents want. This goal is determined by a host of environmental factors that modify the relative attractiveness of many versus few children. Second is the incidence of death, mainly among offspring, which necessitates a compensating adjustment in birth rates to achieve any specific family size goal. Third is the effect of uncertainty in the family formation process where births, deaths, and remarriage are unpredictable.

The family planning hypothesis implies that these factors determine the average level of preferred birth rates, and exert a systematic effect on actual birth rates in following periods. Empirical evidence for Puerto Rico from the 1890s and 1950s is used to test some implications of the family planning hypothesis by juxtaposing observed birth rates and the specified environmental variables or their proxies. The linear multivariate association between local birth rates and the exogenous variables is consistent with the implications of the hypothesis, and in the modern period this association accounts for about half of the variations in birth rates among the 75 municipalities.

Additional factors that are not accounted for by the family planning model are also considered as sources of interregional variation in birth rates. The degree of urbanization and the importance of agricultural activity in the municipality are investigated, but neither variable is significantly associated with birth rates when the other variables of the family planning model are also considered. Because the demographic composition of the population is often linked to differences in fertility, two variables -- age-sex composition and marital status -- are designed to measure the effect of differences in the composition of municipal population on birth rates. Neither of these demographic variables was useful in accounting for the unexplained interregional variation in birth rates.

The general approach to the study of population growth embodied in the family planning hypothesis receives substantial support from this analysis of contemporary Puerto Rican data. But these findings are far from conclusive, and many questions remain unanswered or in tentative form, such as the selection of empirical proxies for unobserved conceptual variables, the specification of the dynamic statistical model and the lag structure it implies, and the system of behavioral equations within which this family planning model should ultimately be imbedded to estimate properly all of the systems parameters.

PN-5405

78

CONTENTS

PREFACE iii

SUMMARY v

LIST OF TABLES ix

Section
I. INTRODUCTION 1

II. A CONCEPTUAL FRAMEWORK 2
The Family Size Goal 2
The Incidence of Death 11
Uncertainty 13

III. TOWARD AN EMPIRICAL MODEL 15
A Birth Function 15
Aggregation 16
Anticipations and Dynamics 18

IV. EVIDENCE ON THE FAMILY PLANNING HYPOTHESIS 23
Introduction 23
The Demographic and Economic Data 23
Premodern Equilibrium: The 1890s 27
Modern Transition: The 1950s 35

V. SUMMARY AND CONCLUSIONS 52
The Hypothesis 52
The Evidence 52
Analytical Shortcomings 54

Appendix
A. ASPECTS OF PUERTO RICAN CASE STUDY 57
B. THE TREATMENT OF UNCERTAINTY IN THE FAMILY FORMATION PROCESS 65
C. TABLES 68
D. ESTIMATION PROCEDURE FOR DYNAMIC RELATIONSHIP BASED ON A COMBINATION OF TIME SERIES AND CROSS SECTION DATA . 71

LIST OF TABLES

1. Definitions of Variables Used in Regression Analysis of Puerto Rico: 1894-1897 28

2. Regressions on Puerto Rican Municipality Birth Rates, 1894-1897 30

3. Final Regressions on Original Variables: Birth Rates in Puerto Rico, 1894-1897 32

4. Regressions on Transformed Variables: Birth Rates in Puerto Rico, 1894-1897 34

5. Definitions of Variables Used in Regression Analysis of Puerto Rico, 1951-1957 37

6. Preliminary Regressions on Municipal Birth Rates and Death Rates for Puerto Rico, 1951-1957 38

7. Preliminary Regressions on Original Variables: Birth Rates in Puerto Rico, 1950-1957 40

8. Final Regressions on Original Variables: Birth Rates in Puerto Rico, 1950-1957 41

9. Regressions on Transformed Variables: Birth Rates in Puerto Rico, 1950-1957 46

10. Subperiod Regression Results: Birth Rates in Puerto Rico, 1950s 48

11. Elasticity of Municipal Birth Rates with Respect to Exogenous Variables: Puerto Rico Selected Periods . . . 49

Appendix

A-1 Labor Force Participation Rates, Number of Children Ever Born, and Median Income of Women in 1960: San Juan, Puerto Rico, and the United States 59

A-2 Ratio of Female to Male Median Incomes by Age and Weeks Worked in 1959: Puerto Rico and United States 62

A-3 Age-specific, Crude, and Standardized Birth Rates per Thousand in Puerto Rico for Selected Years 64

C-1 Variables in Early Regression: 1894-1897 68

C-2 Variables in Modern Regressions: 1951-1957 69

I. INTRODUCTION

This Memorandum explores the hypothesis that parents confronted by similar economic and demographic opportunities and constraints tend to want and to have a similar number of surviving children. Conversely, this working hypothesis implies that differences in reproductive behavior can be attributed to differences in environment. Section II sets forth the conceptual framework for a family planning model based on this hypothesis and discusses its empirical implications. Section III translates the analytical framework into a statistical description of the determinants of regional birth rates. Some of the inferences drawn from the model are then tested in Section IV with the aid of cross-sectional data for Puerto Rico. The concluding section summarizes the empirical findings and raises some issues for further study.

II. A CONCEPTUAL FRAMEWORK

The number of children parents want is a function of their subjective appraisal of the benefits and costs, both psychic and pecuniary, associated with enlarging their family size. A number of characteristics of an environment can be identified that are likely to affect the balance between benefits and costs of children, and thereby alter the number of surviving children parents want. But this desired number of surviving children does not uniquely determine the number of births needed to achieve it. This will depend, as well, on the death rate for children. In addition, decisions with respect to a desired birth rate may be influenced by the uncertainty that surrounds family births, deaths, and remarriage. These three components of the family planning model, the family size goal, the incidence of death, and uncertainty, are treated in turn in this section.

THE FAMILY SIZE GOAL

Parents value their children for themselves and in many cases as a source of income. But balanced against the benefits of having children there are also costs; a child entails two sorts of costs. First, there is the opportunity cost of time parents spend with their children, which is subjectively valued by them relative to their alternative opportunities.[1] Second, there are pecuniary costs of goods and services required to feed, clothe, shelter, and educate a child. Though the combined value of these inputs of child rearing is not entirely independent of the economic resources at the disposal of the parents, income, in terms of time and money, is likely to constrain the number of children parents want.

It is useful to list and analyze those characteristics of a community that may systematically affect the subjective or pecuniary net

[1] The opportunity cost of a child is, more precisely, the difference between the parents' subjective evaluation of spending time with an additional child, which may be positive even in large families, versus the net pecuniary gain and subjective satisfaction associated with devoting their time and energy to other activities.

cost of having children. For pragmatic reasons, empirically observed variables are treated that correspond only approximately to the conceptual variables called for by the analytical framework. Some seemingly important aspects of a community environment, moreover, appear to alter both the costs and benefits of rearing children. In these cases, their net effects on parents' preferences for children cannot be predicted with assurance.

Opportunity Income of Women

To devote her time to her children, a mother forgoes the opportunity to earn additional income or undertake other activities; this opportunity cost of children is an important part of the total costs of rearing children, and a part that may grow as a society advances economically.[1] When women can easily get good jobs outside of the home, they tend both to participate more frequently in the labor force and to appreciate more fully the opportunity costs associated with enlarging their family or lengthening their years of child rearing. One expects, therefore, to find in an environment where women can earn greater incomes (per unit time), higher female participation rates, somewhat lower birth rates, and shorter intervals between births, other things being equal.[2]

Since women often enter the labor force for only part of the year, or seek only part-time employment, annual earnings are not necessarily a good indicator of their opportunity income per unit time worked. For

---

[1] Income opportunities open to men and women are likely to influence a couple's preferred pattern of participation and fertility. The timing of migration and marriage may also be related to the relative strength of demand for the services of men and women in various regional labor markets. Simultaneous treatment of these spheres of behavior is not attempted here, though the specification and estimation of this system of relationships in which these behavioral decisions are endogenously determined will be attempted in a later Memorandum.

[2] Each of these simple associations between women's earnings, activity, and fertility can be verified with U.S. data, and international comparisons of national data also appear to be consistent with these direct implications of the opportunity income hypothesis. Some empirical evidence on these relationships in the United States and Puerto Rico are reported in Appendix A.

men, conversely, who usually seek full-time work from age 20 to 60, annual earnings are a much better measure of earnings per unit of time in the labor force (employed and unemployed). Because of this difficulty in interpreting a woman's annual earnings or income, participation rates are a useful proxy for the opportunity income of women and their access to the labor market.[1]

In addition to increasing the cost of a mother's child care, enhancing her opportunity income status relative to men contributes, within the family unit, to other more complex social developments that may further augment the cost of rearing children. To the extent that a woman becomes a breadwinner in the family, she grows less economically dependent on a man, and less in need of the security provided by a conventional "legal" marriage. In this situation, the man tends to carry fewer of the traditional responsibilities of parent, perhaps because he feels he cannot adequately fulfill the function of provider. Among poor, often urban classes, in both developed and less-developed countries this "unconventional" scheme of social organization is found, where male unemployment is high, female participation is high, unstable common-law union is frequent, and responsibility and authority within the family rests largely on the woman.[2] Without the overhead of the extended family or the provision of welfare for support of dependents, this unconventional family structure raises the net costs of children and child care. The relative income status of men and women is one

---

[1]The need to standardize data on women's annual income for time worked is evident from Table A-2. Holding constant weeks worked, the median income of women is approximately on a par with men in Puerto Rico in 1959, but for all women (and men) with income, their income is only 43 percent that of men. Standardization of income data for weeks worked from published sources is possible in Puerto Rico only at the national level.

[2]Oscar Lewis attributed each of these characteristics to the immigrant communities he studied in Mexico City in 1951: "a higher proportion of gainfully employed in the total population, including child labor and working women, a higher incidence of free unions or consensual marriages, ... a strong tendency toward matricentered families in which the mother played the crucial role...." M. Hanser and L. F. Schnore (eds.), "The Folk Urban Ideal Types," The Study of Urbanization, John Wiley & Sons, New York, 1965, pp. 495-499.

ingredient, but perhaps a crucial one, in this special environment that fosters a matriarchal family order, with its repercussions on child costs.[1]

Child Labor

Children may also be gainfully employed. Until children can earn more than they consume, they are clearly economically dependent on the family. The extent to which these potential child earnings are realized by the parent probably depends on the alternative opportunities open to children assisting their parents or attending school, and the social attitudes toward child labor practices within and outside the home. At a later age when children can earn more than they consume, a variety of cultural and economic factors are likely to determine what fraction of this "net income" is claimed by parents. Regardless of the underlying determinants of child labor practices in the community, the prevalence of unpaid family workers should be associated with lower net costs for rearing children.

Family Income

Typically the main source of family income is that of the male head of household. A change in his income may have a variety of effects on parents' desires for children, depending on whether the change in income is anticipated and whether it is permanent. Because children represent a long-term commitment, parents are not likely to respond to a change in income by adjusting the final number of children they want unless they view the change as permanent. The timing of births, on

---

[1]Greater equality in income opportunities between man and woman has attracted mention but rarely received much emphasis in descriptive accounts of this social setting. Oscar Lewis' "culture of poverty" derives its inspiration from Mexico, Brazil, and Puerto Rico, but parallels between it and Daniel Moynihan's analysis of the plight of the American Negro are clear. Studies of Peru, Colombia, Jamaica, and Puerto Rico have also treated pieces of this social structure, though without explicit consideration of the relative income earning abilities of men and women. Data for Puerto Rico are available bearing on this hypothesis and are discussed in Appendix A. See Table A-2.

the other hand, might be altered in response to transitory changes in income.[1]

An unanticipated but permanent change in family income would appear to change, in the same direction, the number of children parents can expect to rear at their current standard of living. However, in the long run, a permanent shift in income is usually translated into a new standard of living for parents and children alike. Though parents have some latitude in the standard of living they provide their children, social pressures play a prominent role in their choice, and the standards deemed acceptable rise with parents' income and status.[2] Exactly how these countervailing effects of a permanent change in income and a subsequent change in child costs balance out remains uncertain.[3] For

---

[1] Since the empirical portion of this study does not deal with age-specific data on total number of births or birth rates, inferences drawn from this line of reasoning on the interplay between permanent and transitory income and the number and timing of births cannot be tested.

[2] The effect of various "factors mediating the influence of current income on fertility" has been carefully explored in a study of a sample of Detroit women by Ronald Freedmand and Lolagene Coombs, "Economic Considerations and Family Growth Decisions," Population Studies, Vol. 20, No. 2, November 1966. Some other students of the subject have also sought to measure the effect of permanent income (as opposed to current) and relative income (as opposed to absolute) on fertility of U.S. couples. The estimated effect of permanent and relative income on fertility is greater than that of current absolute income, but it would appear that the spacing of births is more affected than their total number. See Deborah S. Freedman, "The Relation of Economic Status to Fertility," American Economic Review, Vol. 53, No. 3, June 1963; and Jay W. Kenvin, "Permanent Income and Child Spacing," preliminary paper given at the meetings of the Population Association of America, Cincinnati, Ohio, April 28-29, 1967.

[3] If the costs of rearing a child were invariant with respect to the parents' income, a permanent change in parental income would probably have a direct effect on the number of children a parent wants. This presumes that children are a superior good, for which superior substitutes are not available. Increased consumption of other goods subsequent to a rise in permanent income diminishes their marginal utility and induces a shift of resources to child rearing. However, the effect of an increase in income is likely also to raise the direct and indirect costs of children as much or more than other goods. Indirect costs refer here to the time inputs of the consumption activity that are evaluated at their shadow price or opportunity value. Though the income effect is probably positive, the price effect is certainly negative, and

empirical purposes, interregional differences in median income levels should correspond closely with differences in permanent income levels, but unfortunately data on personal income are subject to more severe errors of measurement than most other environmental variables.[1]

Education

School for children, even when provided free by the state, imposes opportunity costs on parents. Even if children do not work outside the home, they help out in the home by tending younger children and performing routine household chores.[2] This help is reduced when they attend school. School attendance probably also adds to direct household

---

the net effect of the two is uncertain. Thus, from the point of view of consumer demand theory, a permanent rise in income without holding prices (child costs) constant may have either a positive or negative net effect on parents' desires for a number of children.

The pioneering application of demand theory to human fertility on which this paper relies heavily is by Gary Becker, "An Economic Analysis of Fertility," Demographic and Economic Change in Developed Countries, National Bureau of Economic Research, Princeton University Press, Princeton, New Jersey, 1960. In this paper Becker proposes the distinction between quality and quantity of children demanded by parents. The distinction has been given analytical and empirical content in a study of consumer demand for categories of goods for which quality and quantity dimensions could be separated. (H. Theil, "Quality, Prices, and Budget Enquiries," Review of Economic Studies, XIX(3), No. 50, 1952-53, pp. 129-147.) But this concept applied to the demand for children ignores the important extent to which quality of children is determined by social status and income of parents, as alluded to by James Dusenberry in his comment on Becker's paper, in the same conference volume.

[1]Personal income data are weak for reasons of conceptual scope and systematic response errors. In the first case, income data usually omit non-money income or income in kind, which constitute a substantial fraction of total personal income in a less developed rural-agricultural economy. In the second case, errors in response tend to be different for different types of personal income, which may in turn be related to behavior patterns studied here. For example, entrepreneurial and property incomes are underreported by a larger margin in the United States surveys and censuses than are wage and salary incomes.

[2]Before children can supplement family money income through work outside the home, they provide a variety of services that lighten the tasks of the mother and reduce the costs to her of having additional children. School attendance also broadens the child's social contacts and extends his interests beyond the home, which, according to Oscar Lewis' description of Tepoztlan, tend to further reduce the child's

outlays for better clothes, school materials, transportation and support away from home. The parents' decision to send their children to school increases child costs and where resources are limited reduces the family size goal. The choice between having more children or providing fewer with additional opportunities for education may be an important watershed in the transition from so-called "traditional values" where reproductive behavior is initially consistent with a regime of high childhood mortality and low social mobility, to so-called "modern values" where reproductive behavior has adjusted to conditions of low mortality and greater mobility in accord with the individual's talents, training, and formal education.

The schooling of the parent is also likely to affect the family size goal. Education of parents may influence the amount they are willing to invest in their child's education, and thereby alter their family size goal. Moreover, since education is a good predictor of opportunity income, parents' schooling reflects the opportunity costs of the parents' time spent attending to the needs of their children. Also, schooling may increase the parents' exposure to better methods of birth control, contributing to earlier and more reliable family limitation. Finally, schooling may foster flexibility in behavioral patterns, helping parents cope more quickly with environmental changes, such as a sudden decline in death rates.

Institutions

The form of the family may facilitate income transfers between generations and from active workers to inactive aged and infirm. The extended family tends to shelter both the old and young; parents expect some claim on their offspring's future earnings if their own means of support are exhausted, just as young parents turn to the extended family for support of their children when their current earnings are depressed. The scope and necessity of these family production and

---

contribution to family well being. Oscar Lewis, Life in a Mexican Village: Tepoztlan Restudied, University of Illinois Press, Urbana, Illinois, 1951, pp. 388-389.

redistribution relationships appear to be a function of the average level of wealth in the society and the specific activities undertaken by the state.[1] The government may, for example, proscribe child labor, enforce school attendance, institute support programs for the aged and unproductive, and provide compulsory insurance programs for disability, medical care, and retirement. Public policies further impinge directly on the costs and benefits of children versus other private resource uses by government tax and expenditure policies, which extend, on one hand, personal deductions and dependency allowances on taxes, and on the other hand provide for public health, education, and welfare services.[2]

Contraception

The costs of birth control consist of first acquiring and evaluating information about alternative contraceptive methods and then outlays and inconvenience associated with using a method. Traditional methods of birth control are less reliable and less convenient than modern

---

[1] Little is known about the precise effect of the extended family or nuclear family on member behavior. In particular it is not clear what effect the extended family has on reproductive behavior; even single factor analysis of fertility and family type have not discovered a marked relationship between the two variables. For example, in India minimal difference in age standardized fertility exists between couples living in nuclear and joint family units. Edwin D. Driver, Differential Fertility in Central India, Princeton University Press, Princeton, New Jersey, 1963, Table 69, p. 82. Moreover, it is implicitly argued in this Memorandum that single factor analyses are inadequate to the study of birth rate differences, for many highly correlated factors are at work. The "true" effects of one factor cannot be separately distinguished until all are identified and included in the analysis. Indeed, even then the "true" effects may be intractable to statistical analysis and confident measurement. Liu, in a more thorough and thoughtful study of differences in fertility associated with family structure, cannot identify a consistent effect of the nuclear family structure on fertility compared with the stem or joint family arrangement in regions of Taiwan. Paul K. C. Liu, "Differential Fertility in Taiwan," paper presented at the International Union for the Scientific Study of Population, Sydney Conference, Australia, August 21-26, 1967.

[2] Many of the variables mentioned earlier are also influenced by government policies, but the policy implications that stem from the family planning hypothesis are to be developed in a forthcoming Memorandum.

ones.[1] Where the range of alternatives is limited to traditional methods, large costs must be incurred to achieve a high degree of reliability, as in the extreme cases of continence and induced abortion. For the individual living in a "traditional" community, it may be very costly for him to search independently for a more reliable and a more convenient (modern) method of contraception, whereas for a society as a whole, informational costs are perhaps more modest per capita because of economies of scale both in disseminating information and in distributing the necessary goods and services. When contraceptive supplies are once available and understood, their price will influence their use. Parents, balancing the benefits of more reliable birth control against the costs (including the price) will decide on the amount of supplies they will use in the same manner as they determine their demand for

---

[1] Pregnancy rates differ between populations using the same contraceptive techniques, but these rates probably encompass both developed and less-developed countries, and suggest the differences in reliability between alternative methods. It should be noted that for the U.S. population, Westoff, Potter and Sagi, in The Third Child, Princeton University Press, Princeton, New Jersey, 1963, have shown that pregnancy (failure) rates are relatively high for couples practicing family planning until they have a number of children they want, and thereafter rates are very low. This suggests that failure rates regardless of method are greatly influenced by the couple's motivations.

Type of Method	Contraceptive Technique	Pregnancy Rate per 100 Years of Couple Use
	Douche	40+
	Rhythm or Safe Period	35-40
	Foam tablets	20-40
Traditional	Withdrawal	10-30
	Diaphragm	5-15
	Condom	5-15
Modern	IUCD (IntraUterine Contraceptive Device)	1-5
	Oral Steriod (Pill)	1-5

Sources: Christopher Tietze, M.D., "Intra-Uterine Contraceptive Rings: History and Statistical Appraisal," Intra-Uterine Contraceptive Devices, proceedings of conference April 30-May 1, New York City, Excerpta Medica Foundation, 1962; C. Tietze, "The Use-Effectiveness of Contraceptive Methods," Research in Family Planning, C. V. Kiser (ed.), Princeton University Press, Princeton, N.J., 1962, p. 367, Table 6; "Intra-Uterine Contraception: Research Report," Studies in Family Planning, No. 12, June 1966, pp. 10-11; and Population Council, Population Planning Program, June 1963, pp. 57-58.

other goods.[1] It should be stressed, however, that the personal costs associated with contraception are poorly understood and very difficult to appraise empirically at this time.

Still other factors undoubtedly play a role in determining the relative attractiveness to parents of having few or many surviving children, but for the scope of this exploratory study this abbreviated list of more important factors provides a useful starting point for evaluating the predictive power of the family planning hypothesis. It is assumed, of course, that the environmental variables that are omitted from the analysis are uncorrelated with those included, though this is a shaky supposition. Having considered the effect of these salient features of the parents' environment on their selection of the number of surviving children they want, let us turn to the task of relating that family size goal to the number of births parents prefer.

THE INCIDENCE OF DEATH

In exercising some control over births to achieve a certain surviving size of family, parents may take into account the incidence of death among their offspring.[2] Neglecting for the moment the uncertainty of outcome that stems from the unpredictability of births and deaths

---

[1] If the individual maximizes expected utility, the introduction of a preferred method of birth control yields a benefit equal to the reduction in expected costs (or regrets) of outcomes, minus the added cost or inconvenience associated with adopting the new as compared with the old method. Enough time is allowed for complete adjustment of behavior to the two regimes of birth control.

[2] Parents may also perceive the effect of changes in mortality and health (inducing premature sterility) on the number of years they may expect to bear children, and seek to compensate for this change by earlier or later marriage and childbearing. The extent and ease of remarriage could modify the effect of this factor. However, since the largest percentage declines in death rates occur for the very young and for adults beyond the childbearing ages, I conjecture that probable changes in the incidence of child death are of greater quantitative importance than those of parental death and morbidity on the number and timing of births parents want, other things being equal. Though it would be preferable to separate these effects, both are later combined and treated together in the empirical analysis of crude death rates.

within a particular family, the family planning hypothesis implies that parents try to compensate for the average incidence of death by seeking the number of births that yield the desired number of surviving children.[1] Two mechanisms may be involved in this compensating adjustment of birth rates to death rates. First, the established regime of childhood mortality may influence parents in planning their lifetime reproductive behavior to compensate for what they expect to be the incidence of death among their children. If the death rate of children is high, this adjustment may take the form of earlier marriages or earlier initiation of child bearing. Second, since mortality in childhood is concentrated in the first years of life, parents may make an added effort to have an additional child when they lose one of their previously existing children.[2]

---

[1]Several recent studies present evidence that both at an individual family level and at the aggregate regional level the frequency of deaths appears to influence the subsequent frequency of births. Frederiksen shows that in Ceylon, Mauritius, and British Guiana relative change in birth rates is statistically associated with rates of population growth in the immediately preceding periods which are a reflection of the impact of declines in death rates following the Second World War. He concludes that the improved economic and health conditions that reduced death rates led to the subsequent reduction in birth rates, re-establishing equilibrium rates of population growth. Hassan shows that among sampled Egyptian mothers between the ages of 45 and 49 the loss of a child is associated with having more births thereafter and desiring a larger number of surviving children. These relationships hold for five educational classes. H. Fredericksen, "Determinants and Consequences of Mortality and Fertility Trends," Public Health Reports, Vol. 81, pp. 715-727, U.S. Public Health Service, Washington, D.C., 1966; S. Hassan, "Influence of Child Mortality on Fertility," paper presented at annual meeting of the Population Association of America, New York City, April 1966.

[2]Between the time a decision is made to have a child and the consequent live birth, a series of essentially stochastic events must occur. A period of postpartum sterility (amenorrhea) must pass; conception must occur that comes to term in a live birth. Obviously, if the conception is abortive or stillborn, the entire series of events must be repeated though the period of amenorrhea is somewhat shorter for pregnancy wastage than for completed delivery. In several developed countries evidence indicates that in the absence of contraception, married women in the more fertile ages require a mean interval of some 20 months to have a live birth. In Delhi, India, where lactation is generally prolonged and higher levels of pregnancy wastage are presumed

Although these two behavioral mechanisms are distinct, it is empirically difficult to distinguish between them with the demographic data used in this Memorandum because the usual expectation model and the replacement model imply similar adjustment equations for the purposes of empirical estimation.[1] Regardless of the relative importance of these two mechanisms, the implications are clear that for a community current birth rates are influenced by recent and expected death rates.

UNCERTAINTY

When parents desire a certain number of surviving children they undoubtedly realize that they cannot assure the outcome they want.

---

to result from poor diet and health, the estimated mean birth interval is about three years. Robert G. Potter, Jr., "Birth Intervals: Structure and Change," Population Studies, Vol. 17, No. 155, 1963, pp. 155-166; M. C. Sheps and E. B. Perrin, "Changes in Birth Rates as a Function of Contraceptive Effectiveness: Some Applications of a Stochastic Model," Journal of the American Public Health Association, Vol. 53, No. 7, July 1963, pp. 1031-1046; Savitri Thapar, "Rates and Intervals Between Births in a Population in Delhi," in Papers Contributed by Indian Authors to The World Population Conference, Belgrade, Yugoslavia, 30 August-10 September, 1965, Office of the Registrar General, India, pp. 139-143.

[1] If individual household data were available for births and deaths over time, the replacement hypothesis could be tested against the implications of the expectation hypothesis. (See S. Hassan, "Influence of Child Mortality on Fertility.") Probably both hypotheses have some validity, but when average data are used for a community on birth and death rates, as in this study, the two hypotheses are unfortunately indistinguishable.

To express the replacement hypothesis in a form where the adjustment function of births to past death rates takes a convenient form, it is assumed that the length of the decision-conception-gestation lag is a linear function of a large number of independent population characteristics each of which is normally distributed. Then, the size distribution of lag lengths will tend toward normality as the number of independent characteristic effects is increased. Where the size distribution of lag length is normal about the mean lag length, the cumulative effect on the aggregated behavioral variable (average birth rate) will approach a logistic curve, and the lag structure can be described approximately as one having geometrically declining weights (in prior death rates).

However, the normal distribution is not quite appropriate for the size distribution of the length of lag, since it clearly has a positive

Rather, their course of action influences only the range and probability of expected outcomes. This recognition of uncertainty in the family formation process may induce parents to seek more or fewer children (births) than they would seek under a predictable regime of death and birth rates.[1] For example, where parents emphasize having at least a particular number of children and do not regard additional children as a large liability, uncertainty will tend to raise the desired birth rate. Conversely, where parents regard having too many children as a burdensome demand on their resources and time and regard too few as only a modest inconvenience, uncertainty will tend to depress the desired birth rate, other things being equal. Uncertainty and its effect on the birth rate would be greater under a regime of high death rates and unreliable contraceptive practices than under a regime of low death rates and reliable contraceptive practices.[2] The behavioral and biological assumptions underlying this characterization of uncertainty are set forth more fully in Appendix B.

In summary, parents' decision to seek a particular number of births has been interpreted as a function of (1) the character of their environment that alters their revealed preferences for surviving children, (2) past or expected child death rates that necessitate compensating adjustment in birth rates, and (3) the uncertainty associated with births and deaths that may induce parents to increase or reduce the number of births they seek. This family planning hypothesis must now be translated into a model that yields implications for reproductive behavior that are subject to empirical verification.

---

lower bound and no similar symmetric upper bound. A skewed distribution, such as the lognormal, Pascal, or gamma, might be more appropriate. Experimentation with composite lags in death rates yield more satisfactory distributions, though they involve the incorporation of additional a priori information.

[1] Although not explicitly dealt with here, divorce and remarriage may also enter importantly into the uncertainties of forming a family in some societies.

[2] It is assumed that the number of children that actually survive is symmetrically distributed with respect to the number desired under the regime of either high death rates and unreliable contraceptive practices or low death rates and reliable contraceptive practices. Some evidence for this assumption is presented in Appendix B.

III. TOWARD AN EMPIRICAL MODEL

Parents may decide on the number of children they want in the context of their environment, but a variety of other decisions bearing on their present and future lives may also be involved. The problem is to specify a relation between the number of births parents want and environmental variables that are not themselves determined simultaneously with the desired number of births. In this exploration of the implications of the family planning hypothesis, it is assumed that the factors determining birth rates are themselves predetermined.[1] Neglecting interactions among family behavioral decisions biases our estimates of the effects of environmental conditions on the frequency of births. But, it is hoped this simultaneous equation bias is not large, and the findings of this partial analysis represent a useful though tentative step toward understanding the complex of household decisions pertaining to participation, migration, and fertility.

A BIRTH FUNCTION

It is postulated that parents of different populations have the same (mean) preferences for a number of surviving children, subject to random differences in individual tastes within such populations. The preferences of a parent for a number of children are modified by environmental conditions that alter the absolute and relative costs and benefits of having children.

$$S_{ir} = f(E_{r1}, E_{r2}, \ldots, E_{rm}; Z_i) \qquad (1)$$

where S_{ir} is a vector of net subjective valuations associated with alternative numbers of surviving children for the i^{th} individual in the r^{th} region which is a function of the vector of costs and benefits that the environmental conditions of the r^{th} region, E_{rj}, $j=1,2,\ldots,m$,

[1] A later study will attempt to specify and estimate the more prominent simultaneous relationships among which will be family decisions relating to the frequency or number of births wanted.

offers to the individual from having children and undertaking alternative activities, and Z_i, which equals the sum of the average population preferences for children Z, and differences of taste, e_i, which are assumed to be randomly distributed in space and time with zero mean and constant variance. The preferred number of births, B^*_{ir}, is a function of the vector of net valuations, S_{ir}; the incidence of mortality, d_r; and the uncertainty associated with family planning, u_r:

$$B^*_{ir} = g(d^*_r, u^*_r; S^*_{ir}) \qquad (2)$$

where the asterisks denote that the variables are anticipated or preferred future values of these variables. The transition from one time path toward a desired number of children and another time path is constrained by a feasible range of birth rates and by time lags. Because it takes time to have children and they die unpredictably, the desired path of family formation may diverge from the actual path, during which time family size is said to be in disequilibrium.

This formulation of the birth model has limited empirical application, for inferences pertain to time series for an individual's socio-economic environment and completed fertility. Panel data of this sort are rare.[1] More common are data on the average frequency of births, deaths, school attendance, and so on for a group of individuals that constitute a class, community, or nation. If the implications of this model are to be widely tested against data, the model must be reformulated to describe the relationship between the average frequency of births in a group and this group's past and current character and environment.

AGGREGATION

It is permissible, given particular assumptions, to aggregate individuals' birth functions and derive an expression for the average

[1] Time series analysis is planned of information from a sample of mothers for whom socio-economic data are combined with complete pregnancy and family histories. Unfortunately it is very difficult

desired birth rate in terms of a constant and the average characteristics of the aggregate group. First, it is assumed that the demographic composition of the groups analyzed are similar and, thus, do not affect the desired birth rate.[1] Second, it is assumed that persons in each group face the same environmental circumstances and form their expectations of the future in like manner. Then the annual desired birth rate per thousand population in year t for the r^{th} region, b^*_{rt}, is a function of the group's average environment and a demographic constant:

$$b^*_{rt} = kB^*_{rt} = \frac{k}{n_r}\sum_{i=1}^{n_r} B^*_{rt} = \frac{k}{n_r}\sum_{i=1}^{n_r} g(\bar{d}^*_{rt}, \bar{u}^*_{rt}, \bar{S}^*_{rt}) \qquad (3)$$

where n_r is the group's population size, $\bar{d}^*_{rt}$, $\bar{u}^*_{rt}$, and $\bar{S}^*_{rt}$ are the average anticipated values of the death, uncertainty, and net cost variables, and k is a constant associated with a particular relative age distribution and preferred age-fertility pattern. If the objective birth function were roughly a linear combination of its arguments, the annual desired birth rate could be approximated by the linear model:

$$b^*_{rt} = kB_1\bar{d}^*_{rt} + kB_2\bar{u}^*_{rt} + k\sum_{j=1}^{m} B_{3j}\bar{E}^*_{rj} + kZ + e_r \qquad (4)$$

where e_r is a disturbance term with zero mean, constant variance, unrelated over time or among regions, and uncorrelated with the other variables, and $\bar{E}^*_{rj}$ is the average expected value of the j^{th} environmental variable in the r^{th} region.

As the environment changes, and expectations adjust to these changes, the expected and desired variables represented above are not observed. The behavioral and biological process by which the flow of births is adjusted to changes in desired number of children is

---

to obtain data on their past socio-economic conditions that are relevant to testing the family planning hypothesis.

[1] This assumption is later relaxed in analyzing materials for Puerto Rico, to see if the age/sex/marital status composition of populations accounts for some of the differences in crude birth rates among these populations.

exceedingly complex, but must be specified if one is to proceed beyond this static statement of unobserved equilibrium toward a first approximation of the simpler dynamic properties of this process of human behavioral adjustment.

ANTICIPATIONS AND DYNAMICS

Disequilibrium is due to incorrect anticipations. A systematic error in anticipations leads parents to behave in accordance with a different state of the world than they find themselves living in. If death rates among children drop secularly, for example, parents may conservatively adjust their anticipations of future death rates or they may be unable to compensate fully for the change, and subsequently there will emerge a margin of disequilibrium within the family that will be reflected in the fact that parents will tend to have more children living to adulthood than they had anticipated or originally wanted. The specifications of how parents evaluate change in their environment and form expectations about the future implies the dynamic properties of the family formation process and gives strict empirical meaning to the notion of disequilibrium.

In anticipating the future, one is guided by past experience, where the recent past plays a more important role than the distant past. It is assumed that anticipations are some weighted average of past experiences, the weights probably declining with time. If one's anticipations were based on p prior values of each variable, equation (4) would contain $p(2+m)$ exogenous variables. Severe problems of multicollinearity among these variables would, even where many degrees of freedom obtain, probably preclude identification of the underlying structure of the weights.

A priori information is needed to select which exogenous variables should be treated as anticipated variables and whether the form of these anticipations can be truncated or standardized. In contrast to the gradual evolution of birth rates and most environmental conditions over time, death rates are subject to greater year to year variation. Furthermore, death rates via the replacement hypothesis may induce a

short run response in birth rates that is not due to long run anticipations but actual loss of children. Birth rates may thus be treated as an approximate function of a distributed lagged adjustment to recent death rates (or their anticipated values) and prior levels of the other environmental variables. This formulation conveniently reduces the number of distributively lagged exogenous variables to one, which is also fortunately the one variable for which reliable annual series are available:

$$b_t = kZ + k\sum_{h=x}^{x-p} B_{1h}\,\bar{d}_{t-h} + kB_2\bar{u}_{t-x} + k\sum_{j=1}^{m} B_{3j}\,\bar{E}_{j,t-x} + kv \qquad (5)$$

where p is the number of prior periods during which current anticipations are sensitive to death rates, and x is the minimum conception-gestation lag that must elapse between the decision to have a birth and its arrival.

Another approach to the expectation formation process assumes that the structure of weights geometrically declines with time similarly for all variables, so that all lag structures are specified by a single parameter.[1] If equation (4) prescribes the desired birth rate in terms of current values of the dependent variables, then it is assumed that a fixed proportion of the divergence between the prior birth rate and desired one is eliminated in each subsequent period. The behavioral adjustment equation is then expressed:

$$b_t - b_{t-1} = \delta(b_t^* - b_{t-1}) \qquad (6)$$

where δ is the proportion of the stock disequilibrium expressed in terms of flows that is eliminated in a period. Combining equations (4) and (6), one obtains a single expression for the observed birth rate in terms of geometrically distributed lagged functions of the exogenous variables:

---

[1] For a general treatment of a variety of distributed lag model formulations, see Marc Nerlove, Distributed Lags and Demand Analysis, Agricultural Handbook No. 141, United States Department of Agriculture, U.S. Government Printing Office, Washington, June 1958.

$$b_t = kZ + kB_1 \sum_{h=x}^{t} \delta(1-\delta)^{t-h} \bar{d}_{t-h} + kB_2 \sum_{h=x}^{t} \delta(1-\delta)^{t-h} \bar{u}_{t-h}$$
$$+ \sum_{j=1}^{m} B_{3j} \sum_{h=x}^{t} \delta(1-\delta)^{t-h} \bar{E}_{j,\,t-h} + u \qquad (7)$$

or its simpler expression in autoregressive form where the dependent variable enters on the right side of the equation lagged one period:

$$b_t = k\delta Z + k\delta B_1 \bar{d}_{t-x} + k\delta B_2 \bar{u}_{t-x} + k\delta \sum B_{3j} \bar{E}_{j,t-x} + (1-\delta) b_{t-x} + \delta u \qquad (8)$$

This specification of the model has an attraction: fewer parameters to estimate. But it also conceals hazards for identification and estimation.[1] First, there is no strong reason to presume that expectations or replacement behavior lead precisely to the geometric form of lag functions, nor can one be certain that every E_j is known and observed, and any omission leads to inconsistent least squares estimates of the parameters. The omission of significant, slowly changing environmental variables will add to the residual explanatory power of the lagged dependent variable, biasing up its regression coefficient and biasing down the estimate of the adjustment coefficient and the implied structural parameters. Consistent estimators of the parameters in equation (8) can be calculated in this case only at a substantial cost in their efficiency.[2] Other formulations of the adjustment model could be used, but will not be investigated here.[3]

---

[1] Zvi Griliches, "A Note on Serial Correlation Bias in Estimates of Distributed Lags," Econometrica, Vol. 29, 1961.

[2] N. Leviatan, "Consistent Estimates of Distributed Lags," International Economic Review, Vol. 4, 1963; and Griliches, "Distributed Lags: A Survey," draft paper, University of Chicago, June 1965 (forthcoming in Econometrica). Better methods are now available with superior small sample properties. Cf. M. Nerlove, Economic Studies Quarterly, to be published in December 1967, Tokyo.

[3] The first differenced form of equation (8) that removes most of the biases in estimation due to the trend effects of omitted environmental variables deserves investigation, but many add other forms of serial correlation. This first difference formulation of the model eliminates the constant term (average preferences), and tends to focus attention on short-run reaction patterns that are clearly identified

The statistical properties of the residual term in equations (7) and (8) determine the characteristics of various estimates of the model. It is plausible that particular variables relevant to the determination of regional birth rates have been omitted, which though relatively constant (serially correlated) in each region over time, differ among regions.[1] As a consequence the disturbance term can be decomposed into two parts: a regional effect and a remainder. Following the discussion of Balestra and Nerlove, the residual is expressed[2]

$$u_{it} = \mu_i + \nu_{it} \tag{9}$$

where μ_i and ν_{it} are assumed to be unrelated and ν_{it} serially uncorrelated in each region and independent among regions. Ordinary least squares in such a case produce inconsistent and biased estimates of the parameters because the lagged endogenous variable is correlated with current values of the residuals.[3]

By pooling cross-sections from various time periods the ratio of the variance of the regional effect and the variance of the composite residual term can be estimated, and used to transform the observation to a form in which generalized least squares yields consistent and

---

within the period of time studied. A compromise in generality that permits computation incorporates a priori information to specify the probable form in which the short run replacement lag in death rates enters the estimation equation, and then the long run adjustment lag for all other variables can be estimated (death rates can, of course, also be included among the environmental variables to which the long run adjustments apply). Refinement of dynamic formulation of the model will be dealt with in a later study in light of the patterns of behavior that emerge from empirical analyses.

[1] An alternative interpretation would attribute the regional effect to errors of measurement that differ from region to region. These regional effects might stem from administrative and institutional factors.

[2] Pietro Balestra and Marc Nerlove, "Pooling Cross-Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas," Econometrica, Vol. 34, No. 3, July 1966.

[3] In the absence of lagged endogenous variables the standard errors and the multiple correlation coefficient are biased. G. S. Watson and E. J. Hannan, "Serial Correlation in Regression Analysis," Biometrika II, Vol. 43, parts 3 and 4, December 1956.

asymptotically unbiased estimates.[1] In the next section two sets of data are analyzed with the aim first of testing some of the implications of the family planning hypothesis, and second, of examining the sensitivity of the estimates to alternative specifications of the dynamic relationship and the treatment of regionally correlated disturbances.

---

[1] Balestra and Nerlove, "Pooling Cross-Section and Time Series Data," p. 593. Three estimates of this ratio, ρ, are computed for this purpose in the next section.

IV. EVIDENCE ON THE FAMILY PLANNING HYPOTHESIS

INTRODUCTION

The purpose of this section is to see if differences in birth rates across regions of Puerto Rico are associated, as predicted by our model, with differences in regional environments, and to determine whether these hypothesized multivariate associations are statistically significant. Other factors are also considered as possible sources of interregional variation in birth rates not accounted for by the family planning model. In conclusion, the implied sensitivity of birth rates and population growth to certain changes in environment are estimated for Puerto Rico, and expressed in terms of response elasticities.

THE DEMOGRAPHIC AND ECONOMIC DATA

The following empirical analysis deals with Puerto Rican data for some seventy municipalities. Two time periods are considered: the 1890s and the 1950s. The earlier period is characterized by a premodern equilibrium and predates observable declines in death and birth rates and does not encompass any major environmental changes. The later period is characterized by demographic and economic transition when death and birth rates are falling and family environment is being subjected to a variety of new pressures. No attempt is made here to survey the substantial literature on fertility and family planning in Puerto Rico.[1]

In most countries these two periods, of premodern equilibrium and modern transition, are poorly documented by regional economic and vital statistics. Though Puerto Rico affords unusually reliable data on the

---

[1] The literature is too extensive and disparate to attempt exhaustive citation. A few sources the reader might consult are as follows: Paul K. Hatt, Backgrounds of Human Fertility in Puerto Rico, Princeton University Press, Princeton, New Jersey, 1952; J. Mayone Stycos, Family and Fertility in Puerto Rico, Columbia University Press, New York, 1955; Reuben Hill, J. Mayone Stycos, and Kurt W. Back, The Family and Population Control, University of North Carolina Press, Chapel Hill, North Carolina, 1959; Melvin M. Tumin with Arnold S. Feldman, Social Class and Social Change in Puerto Rico, Princeton University Press, Princeton, New Jersey, 1961.

later period of transition, it presents data on the earlier period that are seriously marred by underregistration of births, a problem still common in most less-developed countries. Since Puerto Rico may represent a special case of accelerated development in the 1950s, consideration of the earlier period, despite the compromises it necessitates in terms of data, is warranted to judge the appropriateness of the model to populations not widely practicing modern methods of birth control and not subject to dislocating socioeconomic changes.

In the decade before Puerto Rico became a territory of the United States, the population of the island is estimated to have been growing at about 1.4 percent per year.[1] Birth rates were probably in the neighborhood of 45 to 50 and death rates 30 to 35 per thousand inhabitants.[2] Registered death rates averaged 30 per thousand in the

---

[1] Growth of the population of Puerto Rico is estimated as 1.43 percent per year in the decade before the 1899 census and 1.52 in the decade following. Jose L. Janer, "Population Growth in Puerto Rico and its Relation to Time Changes in Vital Statistics," Human Biology, Vol. 17, No. 4, December 1945, Table 5, p. 281.

[2] An international cross-sectional regression study by Bogue and Palmore has estimated from contemporary demographic data the statistical relationships between direct and indirect measures of fertility. Using their intercept and regression coefficients one can estimate the crude birth rate for Puerto Rico in 1899, or thereabouts. First, using the child (0 to 4 years) to population ratio, for which the linear correlation with the crude birth rate is .958, we obtain an estimate of 37.2 births per thousand for Puerto Rico in 1899. This is too low. The underestimate is probably due to the exceptionally high incidence of infant mortality in Puerto Rico at the turn of the century, that is unmatched in the contemporary international data used in the Bogue-Palmore study.

Though some of the other relevant factors for predicting crude birth rates, such as infant mortality rates, are not available in these early years of the municipal level, they can be approximately estimated at a national level. Using information on median age of marriage (25), percent ever married between the ages of 45 and 49 (80), infant mortality (conservatively set at 170 per thousand live births), an adjustment for age composition (1.092), as well as the child (0-4): woman (15-49) ratio, we estimate the crude birth rate as 44.73 per thousand inhabitants. This estimate is consistent with the registered and reasonable death rate (31) and the observed rate of population increase over the preceding decade (1.43 percent per year). A somewhat higher infant mortality rate such as 200 per thousand live births yields an estimated birth rate of 47.85. Donald J. Bogue and James A.

period 1888-1898, in accord with the indirect evidence, but registered birth rates averaged only 28 per thousand, suggesting the magnitude of underregistration of births.[1] Since the island was then mostly rural and agricultural in character, it is plausible that underregistration of births was relatively uniform across municipalities and years, and hence differences in registered birth rates reflect approximately proportionate differences in actual birth rates.[2] Registered birth and death rates are studied for the years 1894-1897, since registrations were disrupted in 1898, owing to the change in administration following the Spanish-American War. The United States War Department conducted a census of Puerto Rico in 1899 and this census provides what appears to be the first trustworthy environmental information for all parts of the island. Data from these two sources, the 1899 census and vital registrations, are combined for 68 municipalities for the first analysis.[3]

Palmore, "Some Empirical and Analytical Relations Among Demographic Fertility Measures with Regressions Models for Fertility Estimation," Demography, Vol. 1, No. 1, 1964, pp. 316-338; in particular for intercept and coefficient values see Table 10, p. 329. Unfortunately the authors did not provide standard errors to their regression coefficients.

It is also interesting to note that Janer estimated indirectly with simpler methods that the birth rate in 1887-1899 was about 45.7 per thousand, which lies between our two final estimates. Janer, "Population Growth in Puerto Rico," Table 5, p. 281.

[1] Report of the Census of Puerto Rico, 1899, War Department, Office Director Census of Puerto Rico, Government Printing Office, Washington, D.C., 1900, pp. 113-116.

[2] Puerto Rico, with its population of nearly one million in 1899, had no large cities. San Juan, the largest, showed a population of 32,048, and the only other towns reporting more than 10,000 inhabitants were Ponce and Mayaguez. The rural small farming population was more evenly spread across the 68 municipalities of Puerto Rico than was the case in Cuba where urbanization and plantation agriculture were further advanced. The proportion of the cultivated area owned by occupants was 91 percent in Puerto Rico compared with 43.5 percent in Cuba. See Report of the Census of Puerto Rico, 1899, pp. 19 and 43.

[3] To convert the number of registered births and deaths to the respective crude rates requires estimation of the local populations in the years 1894 to 1897. Geographically the "municipios" are not always comparable between the censuses of 1887 and 1899, nor are the regional totals in the 1887 census regarded as reliable (Janer, "Population Growth in Puerto Rico," p. 279). Thus interpolated

In the second period studied, data are derived from U.S. Census of Population and vital registration materials supplied by the Department of Health in San Juan. Though undoubtedly some underregistration of births continues in this period, it represents a very small fraction of the total. Only since 1950 are birth and death rates both available by residence of mother and deceased, rather than merely by place of occurrence which could bias upward vital rates in towns and cities with superior medical and hospital facilities.[1] The deficiencies of the census data are numerous but not insuperable. From census data on the composition of the municipalities in 1950 and 1960, estimates are interpolated for the intervening years assuming a constant annual rate of change during the decade. This assumption is plausible for gradually changing features of the population and its environment, such as educational attainment, school attendance rates, and female activity rates, but this assumption may be less adequate in describing the evolution of factors that are subject to cyclical fluctuations, such as personal income.[2] The conceptual continuity of urban versus rural residents in the two censuses has been challenged by some observers, but no way was found to correct for this shortcoming.[3]

---

estimates of "municipio" populations were not attempted for the intervening years. Rather it was assumed that all municipal populations grew from 1894 to 1899 at 1.5 percent per year, which is the best estimate of the national rate of population growth in this period. This may be a serious weakness in these early estimates of vital rates, but one that cannot be remedied.

[1]To convert the number of registered births and deaths (to residents) to the respective crude rates, estimates of the municipal population are interpolated between census years, assuming that the decadal growth (or decline) in the population of each municipality occurred at a constant annual rate.

[2]At a national level per capita income increased steadily during the decade 1949-1959, but it is not possible to tell whether median personal income (our census measure of income) followed a smooth path of geometric growth in each of the island's municipalities. For per capita national income in constant prices, see Fuat A. Andic, Distribution of Family Incomes in Puerto Rico, Institute of Caribbean Studies, University of Puerto Rico, Rio Piedras, Puerto Rico, 1964, Table A.3, p. 161.

[3]Carleton's comments at a Milbank conference point out some of the inadequacies of the rural-urban distinction drawn by the census

The measurement of personal income is subject to all the weaknesses of similar data collected in the United States, and the definition of income as only money income severely limits the usefulness of personal income data for interregional comparisons in such a highly rural-agricultural economy.[1] Yet on the whole, the censuses and vital statistics for Puerto Rico probably represent the best body of statistics for any country at Puerto Rico's level of development, and consequently deserve careful consideration despite remaining weaknesses. The variables used in the following regression analysis are defined in Tables 1 and 5, and their means, standard deviations, and published sources are given in Appendix C.

PREMODERN EQUILIBRIUM: THE 1890s

The family planning hypothesis implies that parents' preferences for births are influenced by the character of their environment, and it is further postulated that these factors have some, although lagged, effect on observed birth rates. Three features of the environment were distinguished in Section II: (1) past and expected child death rates; (2) relative attractiveness of having few or many surviving children, which was interpreted in terms of factors affecting the net costs of bearing and rearing children; and (3) uncertainty of the family formation process. In the 1890s only the effect of death rates and a few environmental factors can be investigated.

Cross-sectional information for 68 municipalities are pooled for four years to yield a sample of 272 observations. To tell if the hypothesized association between birth rates and prior death rates conforms to expectations, ordinary least squares estimates of the

---

in Puerto Rico. Robert O. Carleton, Milbank Memorial Fund Quarterly, Components of Population Change in Latin America, Vol. 43, No. 4, October 1965, Part 2, p. 193.

[1] For an evaluation of United States personal income statistics, including census data applicable to Puerto Rico, see T. Paul Schultz, The Distribution of Personal Income, Subcommittee on Economic Statistics, Joint Economic Committee, 88th Cong., 2nd Sess., December 1965, U.S. Government Printing Office, Washington, D.C.

Table 1

DEFINITIONS OF VARIABLES USED IN REGRESSION ANALYSIS OF PUERTO RICO: 1894-1897

Birth Rates:	Registered births per 1000 inhabitants of municipality, where it is assumed that the population in all municipalities was growing at the overall rate of 1.5 percent per year experienced between the 1889 census and the 1899 census.
Death Rates:	Registered deaths per 1000 inhabitants of municipality, estimated as above.
Education Variables:	
	A. Percent of children between ages of 7 and 9 attending school in 1899.
	B. Percent of children 10 to 17 years of age in school in 1899 (students over 17 included).
	C Percent of persons 10 years and over who were able to read and write in 1899.
Female Activity:	Percent of women 18 years of age and older with "gainful occupation."
Urbanization:	Percent of municipal population living in towns larger than 1000 inhabitants in 1899.
Married Proportion:	Percent of municipal population 18 years of age and older legally or consensually married in 1899.

linear relationships are presented in Table 2.[1] In regressions 1 through 8 we experiment with alternative combinations of prior death rates to estimate directly the form of the distributed lag function, but the high correlation between local death rates in successive years precludes defining the lag structure for more than a few years.[2] However, all statistically significant coefficients are positive and tend to conform to the expected declining pattern.

If the lag weights for prior death rates decline geometrically with time, and no other variables play a role in determining birth rates, the least squares procedure is appropriate for estimating the parameters to the proportional adjustment model. Estimates of the proportional adjustment model are reported in regression 9, Table 2. They imply that in the short run, the birth rate responds about .10 to a unit change in last year's death rate, and in the long run the birth rate responds .74 to a permanent unit change in the death rate.[3] This indirect estimate of the long-run response coefficient exceeds those directly estimated in regressions 1 to 8, which range between .48 and .62.[4]

A variety of environmental variables can be derived from the 1899 Census of Puerto Rico that reflect characteristics of the municipal populations of their environments that probably alter the net-

---

[1] It seems reasonable to assume, particularly in this early period, that regional variations in the crude death rate reflect variations in the child death rate, for from age specific death records beginning in 1910, it is evident that one-half of the deaths struck children less than 15 years old. This proportion was slightly more than one half in 1910 and slightly less than one half by 1940, declining to around a third by the 1950s. Janer, "Population Growth in Puerto Rico," Appendix Tables 5a through 5f.

[2] The simple correlation coefficient for death rates in successive years is approximately .7 in this early period.

[3] The long-run response coefficient is derived by dividing the short-run coefficient for the lagged death rate variable by one minus the coefficient for the lagged birth rate, that is, .097/(1-.868).

[4] If relevant variables are omitted and residuals autocorrelated the least squares estimates of the long run response of birth rates to death rates tends to be biased up.

Table 2

REGRESSIONS ON PUERTO RICAN MUNICIPALITY BIRTH RATES, 1894-1897[a]
(standard errors of regression coefficients shown in parentheses)[b]

Regression Number	Constant Term	Lagged Death Rates (years)					Birth Rate Lagged One Year	R^2
		t-1	t-2	t-3	t-4	t-5		
1	13.29	-	.328[d] (.126)	-.032 (.144)	.231[c] (.119)	.015 (.100)	- -	.118
2	13.35	-	.330[d] (.126)	-.027 (.141)	.239[c] (.106)	-		.118
3	14.37	-	.379[d] (.125)	.137 (.121)	- -	-	-	.101
4	15.37	-	.478[d] (.089)	-	-	-	-	.097
5	11.28	.340[d] (.129)	.141 (.143)	-.085 (.141)	.227[c] (.105)	-	-	.140
6	12.17	.353[d] (.130)	.181 (.143)	.068 (.122)	-	-	-	.125
7	12.56	.368[d] (.127)	.220 (.125)	-	-	-	-	.124
8	14.18	.527[d] (.089)	-	-	-	-	-	.114
9	.586	.097[c] (.047)	-	-	-	-	.868[c] (.030)	.780

Notes:

- Variable not included.

[a]Definitions of variables are given in Table 1.

[b]Number of observations = 272.

[c]Regression coefficient is significant at .05.

[d]Regression coefficient is significant at .01.

cost schedule for parents having children, and may thus affect local birth rates. Since the previous census cannot be linked for the purposes of interpolating yearly estimates, it is assumed that these environmental variables did not appreciably change in the five years 1894 to 1899. The level of education is measured in three ways: school attendance rates for children 7 to 9 and 10 to 17 years of age, and adult literacy. Other variables considered in addition to prior death rates and education are female activity rates, and a measure of urbanization. Table 3 shows the regression results using these variables.

The regression coefficients on all environmental variables taken in combination with death rates are of the expected sign, and all but female activity rates individually satisfy standard tests of statistical significance. School attendance and adult literacy are associated with lower birth rates, though the school attendance rate for children between the ages of 10 and 17 is slightly more powerful in explaining inter-regional variations in birth rates than the other two education variables.

Characterizing an environment as urban or rural neglects more fundamental differences among environments that are only approximately associated with this arbitrary dichotomy. Urban as compared with rural residence frequently involves a number of environmental changes that on balance appear to add to the costs of rearing children, and thus may foster lower birth rates among urban than among rural populations.[1] Pursuing the objective of this Memorandum to investigate how environment influences reproductive behavior, these particular characteristics of the family environment associated with child costs are first considered,

---

[1] A number of factors might be enumerated that could add, on balance, to the real costs of rearing children in the urban compared with the rural environment: (1) direct costs of food and housing are likely to be greater for a large family; (2) opportunities to use the productive talents of children to add to family resources tends to be more limited; (3) the need to invest in one's children's schooling and skills is better understood in the urban environment, and hence the associated costs are more often accepted; (4) opportunities for women to work outside of the home are greater, and consequently their potential contribution to family income that is forgone while children are young is larger.

Table 3

FINAL REGRESSIONS ON ORIGINAL VARIABLES: BIRTH RATES IN PUERTO RICO, 1894-1897[a]
(standard errors of regression coefficients shown in parentheses)[b]

Regression Number	Constant Term	Lagged Death Rates		Education			Urbanization	Female Activity	ρ	R^2
		t-1	t-2	(A)	(B)	(C)				
1	14.25	.390[d] (.125)	.258[c] (.123)	-.465[e] (.138)					.852	.160
2	16.98	.378[d] (.124)	.237 (.122)		-.609[e] (.169)				.854	.165
3	15.50	.411[d] (.125)	.227[c] (.124)			-.359[e] (.102)			.849	.163
4	12.37	.432[c] (.126)	.264[c] (.123)				-.151[e] (.042)		n.a.	.164
5	13.25	.367[d] (.127)	.222 (.125)					-.042 (.060)	n.a.	.126
6	15.11	.411[d] (.126)	.256[c] (.123)		-.366 (.231)		-.088 (.057)		.849	.172
7	17.19	.378[d] (.125)	.238 (.123)		-.604[e] (.171)			-.015 (.059)	.854	.165

Notes:

n.a.: Not available.

[a]Definitions of variables are given in Table 1.

[b]Number of observations = 272.

[c]Regression coefficient is significant at .05.

[d]Regression coefficient is significant at .01.

[e]Regression coefficient is significant at .001.

and urbanization is then introduced as a possibly useful proxy for the residual characteristics of the urban social matrix that defy economic conceptualization or quantitative measurement. In regression 4 the degree of urbanization is negatively associated with birth rates to a statistically significant extent. Yet with the introduction of any of the three education variables, as in regression 6, the coefficient for urbanization declines to one-half its earlier level and is no longer statistically significant.[1]

Female activity rates are weakly associated with lower regional birth rates, as hypothesized, but this association becomes completely insignificant after the inclusion of any of the education variables (regressions 5 and 7).

Since birth registration materials for this period are clearly deficient, one must consider whether the degree of underregistration of births might not be a persistent regional characteristic determined by local institutional conditions beyond the scope of this investigation. In this case the disturbance term in the estimation equation should be represented as a combination of a regional effect and a stochastic remainder (equation (9)). Estimating ρ, the ratio of the variance of the regional effect to that of the composite disturbance term, this ratio is found to equal approximately .85 in the final regressions reported in Table 3.[2]

A generalized least squares procedure based on this estimate of ρ yields new estimates of the parameters that are presented in Table 4. Adjustment of the data for regional effects reduces the size of all coefficients and markedly reduces their statistical asymptotic significance. Though these estimates should be interpreted cautiously because of their relatively large asymptotic standard errors, they are

---

[1] Education (and income) differences between the urban and rural environments seem to be the principal factor accounting for differences in birth rates between these two types of regions, and in the later period jobs for women appear to contribute to the urban-rural fertility differences.

[2] Appendix D shows the generalized least squares procedure used in these estimations.

Table 4

REGRESSIONS ON TRANSFORMED VARIABLES: BIRTH RATES IN PUERTO RICO, 1894-1897[a]

(standard errors of regression coefficients shown in parentheses)[b]

Regression Number	Constant Term	Lagged Death Rates		Education			Urbani-zation	Female Activity	R^2
		t-1	t-2	(A)	(B)	(C)			
1	14.38	.112 (.067)	.010 (.065)	-.322 (.269)					.015
2	15.64	.110 (.067)	.007 (.065)		-.535 (.332)				.019
3	14.69	.113 (.067)	.011 (.066)			-.188 (.194)			.013
4	15.75	.108 (.067)	.006 (.066)		-.595 (.452)		.021 (.108)		.019
5	15.69	.110 (.067)	.007 (.065)		-.533 (.336)			-.007 (.117)	.019

Notes:

[a]Definitions of variables are given in Table 1.

[b]Standard errors are asymptotically unbiased. Numbers of observations = 272.

nevertheless usefully expressed in terms of response elasticities measured at regressions means.[1] Compensating for the average margin of underregistration of births, the second regression in Table 4 implies that a one percent change in birth rates, and a one percent change in the initially low level of school attendance (for children between the ages of 10 and 17), are inversely associated with a .24 percent change in birth rates.[2]

The relative importance of municipal differences in registered birth rates that persist within municipalities over time and are not accounted for by the available environmental proxies precludes a powerful test of the family planning model in the early period. The generalized least squares estimates are substantially different from those estimated directly, and as one would expect, the asymptotic significance and size of the estimates are reduced by eliminating the regional effects. The weakness of the demographic data and the gaps in the environmental data for the 1890s cannot be remedied, but for the 1950s the statistical materials are more reliable and abundant, and satisfactorily test some of the implications of the family planning hypothesis.

MODERN TRANSITION: THE 1950s

Administrative reforms initiated in 1941 coupled with the rapid expansion of war production accelerated Puerto Rico's economic, social, and political development in the early 1940s, and though postwar readjustments posed problems for the island's economy, the pace of growth has fluctuated but not slackened since the war. In the period 1940 to 1960, per capita real income increased at an average annual rate of 4.5 percent. The younger population migrated first from traditional agricultural pursuits to centers of urban activity, and then large numbers continued

---

[1] Response elasticities represent the relative association between various exogenous variables and the endogenous variable, where it is assumed that the other exogenous variables do not change. Regression means are taken on the basis of the untransformed data.

[2] Since births are underregistered by about one third, the estimated regression coefficients are assumed to be approximately two-thirds their true value.

to the mainland United States.[1] The commonwealth association with a larger and wealthier factor market conferred on Puerto Rico distinct advantages: the factor imbalances that often arise and constrain the course of development were much relieved, first by the establishment of a tax haven to attract needed capital inflows, and second by the provision of an outlet for migration to absorb a part of the rapid increase in the urban labor force. As a consequence and cause of this pattern of growth and population movement, death rates fell abruptly from 19 per thousand inhabitants in 1941 to 10 in 1950 and 7 in 1960. Infant mortality declined comparably from 116 per thousand births in 1941 to 68 in 1950 to 44 in 1960. Birth rates behaved in a more complex fashion; initially they rose from their low in the depressed 1930s to reach a high of 42 per thousand inhabitants in 1945-47. Since then birth rates have gradually declined to 39 in 1950, 33 in 1960, and 30 per thousand inhabitants in 1965.[2]

The period analyzed here is 1951 to 1957.[3] The transition of death rates is nearing its end, falling from 10 to 7 per thousand, and birth rates are beginning to decline, falling from 38 to 34 per thousand inhabitants. Regressions 1 through 4 in Table 6 show the linear

---

[1] See the review of population movement in Puerto Rico by Irene B. Taeuber, "Migration and Transformation: Spanish Surname Population and Puerto Ricans," Population Index, January 1966, pp. 3-34; and for annual estimates of migration flow to the mainland, Stanley Friedlander, Labor Migration and Economic Growth, MIT Press, Cambridge, Massachusetts, 1965, Table 3.1, p. 47.

[2] Some part of the decline in the crude birth rate has been due to changes in the age structure of the population, which in turn may have been fostered by the migration to the United States. However, migration has received too much credit for the decline in crude birth rates by many observers. Table A-3 of the Appendix shows that if the age composition of the population is held constant with 1950 weights, the age-standardized birth rate falls nearly as much as does the crude birth rate until 1958. Migration, according to Friedlander's estimates, reached its peak in 1953, and virtually came to an end after 1959. Thus, the effect of migration in the 1950s on the age composition did not lead to an appreciable divergence between crude and age-standardized birth rate. Since 1958 the age-standardized birth rate has fallen more slowly than the crude rate.

[3] Death rates for the prior seven years, used in calculating the uncertainty variable, include 1943-1956. All environmental variables are lagged one year, and death rates are lagged various periods.

Table 5

DEFINITIONS OF VARIABLES USED IN REGRESSION ANALYSIS OF PUERTO RICO, 1951-1957

Birth Rates: Registered births by residence of mother per 1000 inhabitants of municipality, where it is assumed that the municipality population grew at a constant exponential rate between censuses.

Death Rates: Registered deaths by residence per 1000 inhabitants, estimated as above.

Education Variables:

A. Percent of children between ages of 7 and 13 attending school.[a]

B. Percent of children between ages of 14 and 17 attending school.[a]

C. Median years of schooling completed for all persons 25 years of age and older.[a]

Female Activity: Percent of women 14 years of age and older that are in the civilian labor force.[a]

Unpaid Family Workers: Percent of unpaid family workers in civilian labor force.[a]

Income: Median income of person 14 years of age and older, in current dollars.[b]

Urbanization: Percent of municipal population living in towns larger than 2500.[a]

Agriculture: Percent of civilian labor force in agriculture, forestry, and the fishing industry.[a]

Married Proportion: Percent of municipal population 18 years of age and older legally or consensually married.[a]

Uncertainty: Estimated variance of the probability of death over the preceding seven years, with allowance for linear change in average incidence of death over time.[c]

Notes:

[a] Intercensal observations are interpolated, assuming that the relative rate of change is constant between each pair of census observations for each municipality.

[b] Median income for both sexes in 1960 is not published; it is estimated here as a weighted average of male and female median incomes. Interpolation is used to estimate income level for years between the censuses, assuming a constant rate of change during the 1950s.

[c] Derivation of these estimates of uncertainty stemming from death in the preceding seven years is discussed in Appendix B.

Table 6

PRELIMINARY REGRESSIONS ON MUNICIPAL BIRTH RATES AND DEATH RATES FOR PUERTO RICO, 1951-1957[a]

(standard errors of regression coefficients are shown in parentheses)[b]

Regression Number	Constant Term	Lagged Death Rates (Years)				Birth Rate Lagged One Year	R^2
		t-1	t-2	t-3	t-4		
1	23.45	1.196[c] (.123)	-	-	-	-	.152
2	22.85	.839[c] (.214)	.405[c] (.198)	-	-	-	.159
3	22.73	.798[c] (.232)	.356 (.225)	.095 (.207)	-	-	.159
4	22.76	.806[c] (.238)	.365 (.233)	.109 (.229)	-.030 (.198)	-	.159
5	2.85	.013 (.065)	-	-	-	.896[d] (.021)	.809
6	2.57	-	.081 (.058)	-	-	.887[d] (.021)	.810
7	2.44	-	-	.095 (.056)	-	.886[d] (.020)	.810
8	2.96	-	-	-	-.015 (.051)	.900[d] (.021)	.809

Notes:

- Variable not included.

[a]Definitions of variables are in Table 5.

[b]Number of observations = 525.

[c]Regression coefficient is significant at .05.

[d]Regression coefficient is significant at .001.

relationships between birth rates and prior death rates. The lagged association is statistically well defined for two prior years and the regression coefficients are of diminishing positive value for four years. Regressions 5 through 8 in Table 6 are least squares estimates of the proportionate adjustment model, including only various lagged death rates. The geometric lag structure implied by the proportionate adjustment model in this crude form does not receive much support from these data. The directly estimated relationships imply a unit change in death rate is subsequently associated with a 1.20 to 1.25 unit change in birth rates, whereas the best estimates of the proportionate adjustment model (regression 7) imply a unit change in death rates is associated in the long run with a .83 unit change in birth rates. The final model incorporates death rates directly lagged for two preceding years, since beyond this point they do not significantly add to the explanatory power of the model.

The Censuses of Population in 1950 and 1960 yield a richer variety of environmental information than the census conducted in 1899; also one suspects the later sources are more reliable. Three educational variables are investigated measuring school attendance rates for children 7-13, 14-17, and adult median years of schooling. All three are found in Table 7 to be inversely associated with birth rates. Since school attendance rates for children and adult education add to the costs of rearing children via different mechanisms, directly and indirectly, these effects of the educational process on the family formation process are jointly estimated in regression 4 in Table 7.

Female activity in the labor force reflects greater income opportunities for women outside of the home, and consequently higher opportunity costs for rearing children. In the opposite case, where family members work in the family without pay, some of the costs of rearing children are offset by their productive contribution to family resources. These two reflections of other costs entering into the family formation process are recorded in the censuses, and their inclusion in regression 1 of Table 8 yields our estimates of the complete family planning model. All coefficients are statistically significant and of the appropriate sign.

Table 7

PRELIMINARY REGRESSIONS ON ORIGINAL VARIABLES: BIRTH RATES IN PUERTO RICO, 1950-1957[a]

(standard errors of regression coefficients shown in parentheses)[b]

Regression Number	Constant Term	Lagged Death Rate		Education			R^2
		t-1	t-2	(A)	(B)	(C)	
1	46.73	.789[d] (.205)	.095 (.195)	.789[d] (.205)	-	-	.232
2	33.16	.827[d] (.206)	.288 (.192)	-	.194[d] (.030)	-	.220
3	35.22	.765[d] (.192)	.367[c] (.177)	-	-	-3.273[d] (.287)	.327
4	48.08	.744[d] (.188)	.186 (.179)	-.174[d] (.040)	-	-2.885[d] (.295)	.351

Notes:

- Variable not included.

[a]Definitions of variables are in Table 5.

[b]Number of observations = 525.

[c]Regression coefficient is significant at .01.

[d]Regression coefficient is significant at .001.

Table 8

FINAL REGRESSIONS ON ORIGINAL VARIABLES: BIRTH RATES IN PUERTO RICO, 1950-1957[a]

(standard errors of regression coefficients shown in parentheses)[b]

Regression Number	Constant Term	Lagged Death Rate		Education		Female Activity	Family Workers	Income	Urbanization	Agriculture	Married Proportion	Uncertainty	ρ	R^2
		t-1	t-2	(A)	(C)									
1	36.15	.860[e] (.173)	.324[c] (.164)	-.088[c] (.037)	-1.575[e] (.300)	-.178[e] (.030)	.749[e] (.093)	-	-	-	-	-	.126	.460
2	36.16	.808[e] (.174)	.288 (.165)	-.079[c] (.037)	-.982[e] (.176)	-.176[e] (.030)	.703[e] (.095)	-.005[c] (.002)	-	-	-	-	.127	.465
3	38.53	.837[e] (.173)	.278 (.167)	-.107[d] (.039)	-1.777[e] (.326)	-.186[e] (.030)	.742[e] (.094)	-	.019 (.012)	-	-	-	.123	.463
4	35.68	.858[e] (.174)	.324[c] (.165)	-.086[c] (.040)	-1.540[e] (.354)	-.177[e] (.031)	.743[e] (.098)	-	-	.004 (.019)	-	-	.126	.460
5	35.82	.858[e] (.176)	.323 (.167)	-.088[c] (.040)	-1.575[e] (.301)	-.178[e] (.030)	.750[e] (.094)	-	-	-	.004 (.088)	-	.126	.460
6	36.15	.882[e] (.175)	.352[c] (.167)	-.093[c] (.038)	-1.528[e] (.304)	-.177[e] (.030)	.742[e] (.094)	-	-	-	-	-.924 (.947)	n.a.	.461

Notes:

- Variable not included.

n.a.: Not available.

[a] Definitions of variables are given in Table 5.

[b] Number of observations = 525.

[c] Regression coefficient is significant at .05.

[d] Regression coefficient is significant at .01.

[e] Regression coefficient is significant at .001.

The treatment of personal income raises difficult problems for conceptual analysis, empirical measurement, and causal interpretation. Although the opportunity income of mother and child have predictable and opposite effects on the costs of rearing children, the opportunity income of the father or the family as a whole does not. Furthermore, census data on personal income in Puerto Rico measure only money income, neglecting home produced and used goods and services or those acquired through local barter arrangements. This neglect of income in kind depresses the measured income compared with real income in rural-agricultural regions.[1] Finally, personal income or earnings are a function of a person's formal education, on-the-job training, general and specific experience and talents, plus an element of chance. Inclusion of adult schooling has already taken account of formal education enhancing the parents' opportunity incomes, and one anticipates that it will be difficult to distinguish confidently between the effects of income and adult education on family resources.

The second regression in Table 8 includes family income in addition to the central variables of the family planning model. Income is inversely related to birth rates, though the magnitude of the effect is small and undoubtedly part of its contribution to the explanation of birth rates is at the expense of the adult education variable (C). Because of the conceptual and statistical ambiguity of this measure of income, it is omitted from later analyses, though the effect of income clearly warrants further investigation.[2]

---

[1] See fn. 1, p. 7.

[2] Several additional facts entered into this decision to dispense with this income variable. First, the bias of a money income concept exaggerating income differences between urban and rural-agricultural regions was confirmed by the fact that when income was added to regressions 3 or 4 in Table 8, the coefficient for urbanization became highly significant and positive and the coefficient for agriculture became highly significant and negative -- both contrary to expectations though consistent with the hypothesized bias implied by the concept of money income. Second, when subperiods were analyzed (see below), the coefficient for income declined in size and its standard error increased. No significant relationship between income and the birth rate persisted into the later subperiod, 1955-1957. As income lost its significance in this later period, the perverse effects of urbanization and agricultural variables also disappeared. Third, the personal income

Two other characteristics of the environment have been linked to differences in birth rates: rural-urban residence and agricultural-nonagricultural livelihoods. Both of these variables can be obtained from the Puerto Rican censuses, but they are not significantly associated with municipal birth rates when added to the core of the family planning model. Although rural and agricultural regions do tend to register higher birth rates, one must conclude that the particular aspects of the rural-agricultural environment that influence birth rates have already been described in terms of the other economic and demographic variables.[1] When proper account is taken of economic and health conditions in the environment, there are no significant differences in birth rates between the city and countryside, or between the factory and farm population.

Further detail on the family environment can be attained from the Puerto Rican material for testing other hypotheses of the family planning model. Most difficult to test is the hypothesis that parents react to the presence of uncertainty in the family planning process by altering their birth rate. Uncertainty involves both the timing and the arrival of family births and the timing and incidence of family deaths. No data are available on the frequency and intervals of births given preferred birth schedules. The frequency of deaths, however, is known for the recent past within each community, and is relatively independent of individual preferences. The variability of death can be estimated from past time series. A modification of this measure of uncertainty allows for the interaction of time and expectations, although in a very simple manner. Since death rates declined

---

series was linked between censuses by an indirect adjustment procedure (see note b, Table 5) and in the case of income the assumed constant rate of geometric growth over the decade between the censuses may not be appropriate. All this evidence leads one to suspect that the income variable is not well specified for our purposes and may seriously distort other estimates.

[1]The simple correlation between municipal birth rates and urbanity (negative) and agricultural occupations (positive) persists after one takes account of differences in mortality, but disappears when education and female and family worker activity rates are added to the regression equation.

markedly in the period under study, 1943 to 1956, as health and economic conditions improved in Puerto Rico, and these improvements were largely irreversible, it is assumed that changes occurring systematically over time are incorporated in future expectations. Consequently, uncertainty in death rates is estimated as the variance from a linear time trend that explicitly allows for environmental changes that occur linearly through time.[1] How well this proxy represents the parents' expectations of the uncertainty they will face in forming their family is moot, but the need to deal explicitly with this potentially important factor warrants investigation.

Regression 6 in Table 8 shows an inverse relationship exists between birth rates and this measure of uncertainty, after allowing for the effects of the other environmental variables, but the coefficient for the uncertainty variable is not quite statistically significant at the .1 level. One may infer that people are not reacting systematically to this measure of uncertainty. Much greater refinement of our conceptual understanding and empirical measurement of uncertainty as it enters into the family planning process will be required before any firm conclusions will emerge as to the causal link between uncertainty and human reproductive behavior.

Thus far the demographic composition of the population has been neglected as a possible determinant of birth rates, but it is plausible that differences in the composition of municipal population might affect birth rates to a noticeable extent. It is therefore postulated that differences in age, sex, and marital status of the population account for some part of the variation in crude birth rates across municipalities. To test this hypothesis, two variables are constructed to measure the effect of these compositional characteristics on birth rates.[2] Adding either or both of these variables to the central

---

[1] Appendix B, which discusses the data sources, sets forth the model underlying this estimation of uncertainty in the family planning process.

[2] As a proxy for the share of the population subject to the risk of child bearing, the proportion of adults married was used. Age-specific marital status data was not available by region for the purposes of deriving a measure of the fertile population subject to the risk of

variables of the family planning model does not add to its explanatory power, nor do these variables show a statistically significant association with birth rates or alter substantially the previous estimates of the model.[1] These demographic variables may not vary sufficiently across regions for us to identify their effect on birth rates, or their effect may have already been accounted for by underlying economic variables that themselves influence marriage rates and through their effect on migration change the age structure. But regardless of the reasons, this simple test finds no support for the hypothesis that these compositional features of the municipal populations are helpful in understanding existing differences in municipal birth rates.

Again the question arises whether the unexplained interregional variation in birth rates is not due to a combination of regional effects and a "true" stochastic effect. Estimating the value of ρ for the final regression in Table 8, one discovers the estimated variance of the regional effect is equal to about .126 of the variance of estimated residuals. Table 9 shows generalized least squares estimates when the data are adjusted for these regional effects. The differences between the coefficients in Tables 8 and 9 are not substantial and the asymptotic statistical significance of the new estimates are high by conventional standards for all but the coefficient for the death rate lagged two years. Although regional effects in the 1890s data appeared to bias the direct

---

child bearing. The age/sex proxy was simply taken as the proportion of women in the population in the fertile ages in the 1890s (Table 1), but more detailed data were available for 1950. The women in each age group were attributed the appropriate national age-specific birth rate, and the estimated number of births for each municipality was then divided by the population to derive an estimate of the municipal crude birth rate. This procedure assumes that age-specific birth rates do not vary appreciably across Puerto Rico, so that age/sex compositional differences will account for some of the observed interregional variation in crude birth rates.

[1]The proportion married in both the 1890s and 1950s was positively correlated with birth rates taken alone. However, the more complete model incorporating the central variables in the family planning model removes any residual association between the proportion of adults married and the birth rate. The age compositional variable was not systematically related to birth rates in either period, with or without the addition of the other economic and demographic variables.

Table 9

REGRESSIONS ON TRANSFORMED VARIABLES: BIRTH RATES IN PUERTO RICO, 1950-1957[a]

(standard errors of regression coefficients shown in parentheses)[b]

Regression Number	Constant Term	Lagged Death Rate		Education		Female Activity	Family Workers	Income	Urbanization	Agriculture	Married Proportion	Lagged Birth Rate	R^2
		t-1	t-2	(A)	(C)								
1	26.88	.835[d] (.171)	.304 (.163)	-.083[c] (.037)	-1.451[d] (.296)	-.179[d] (.030)	.766[d] (.095)	-	-	-	-	-	.449
2	26.86	.780[d] (.172)	.268 (.163)	-.072 (.037)	-.834[c] (.394)	-.177[d] (.030)	.719[d] (.097)	-.005[c] (.002)	-	-	-	-	.455
3	28.47	.817[d] (.172)	.263 (.166)	-.099[c] (.039)	1.627[d] (.324)	-.185[d] (.030)	.759[d] (.095)	-	.016 (.012)	-	-	-	.451
4	26.20	.831[d] (.172)	.303 (.163)	-.078[c] (.039)	-1.382[d] (.352)	-.177[d] (.030)	.755[d] (.099)	-	-	.007 (.019)	-	-	.449
5	27.57	.839[d] (.174)	.308 (.165)	-.085[c] (.040)	-1.453[d] (.297)	-.179[d] (.030)	.765[d] (.095)	-	-	-	-.012 (.087)	-	.449
6	5.36	-	.099 (.064)	-.040 (.021)	-.062 (.176)	-.011 (.018)	.245[d] (.057)	-	-	-	-	.823[d] (.025)	.817
7	5.43	-	.068 (.066)	-.035 (.022)	.215 (.229)	-.012 (018)	.226[d] (.058)	-.002 (.001)	-	-	-	.818[d] (.025)	.818

Notes:

- Variable not included.

[a]Definitions of variables are given in Table 5.

[b]Number of observations = 525. The standard errors are asymptotically unbiased in regressions 1 through 5, but their properties in regressions 6 and 7 are not yet well established. See fn. 1, p. 47.

[c]Regression coefficient is significant at .05.

[d]Regression coefficient is significant at .001.

estimates, these effects in the 1950s data are more modest and do not change the estimates appreciably. The autoregressive form of the proportionate adjustment model is estimated in regression 6 in Table 9. The coefficients are of appropriate sign and approximate magnitude, but only the lagged birth rate and family worker variable are asymptotically statistically significant.[1] The inclusion of income in the proportionate adjustment model in regression 7 reverses the sign on adult education, confirming again the close association between these two variables in their effect on birth rates. On balance the evidence seems to point to the inadequacy of the geometric lag formulation of the dynamic adjustment model for this problem.

Is the model completely and correctly specified? If some variables are omitted are they correlated with those already included, biasing our estimates? No firm answer can be given to these questions. Additional environmental factors deserve examination in the context of the family planning hypothesis, and undoubtedly some will account for part of the unexplained interregional variation in birth rates. Exactly how the inclusion of these factors will alter the other estimates must await further research.

One possible weakness in the model is that the hypothesized stochastic relationship between environmental variables and birth rates is not invariant over time, either because the functional parameters are changing, or because functional relationships assumed to be linear are not. To investigate this question, regression 1 of Table 8 is reestimated for two subperiods, 1951-1954 and 1955-1957. The estimated values of ρ for the two subperiods are used to transform the data, and generalized least square estimates of the parameters are shown in Table

---

[1] In the case where a lagged endogenous variable is an explanatory variable, these estimates appear to have desirable asymptotic properties. Amemiya has shown they are both asymptotically unbiased and efficient, A Note on the Estimation of Balestra-Nerlove Models, Technical Report No. 4, Institute for Mathematical Studies in the Social Sciences, Stanford University, Stanford, California, August 14, 1967. However, Nerlove is now engaged in Monte Carlo investigations that suggest that these estimates may have less desirable small-sample properties when exogenous variables enter the relationship.

10.[1] For the purposes of comparison, these final estimates for both subperiods and the total period are expressed in Table 11 as response elasticities of birth rates to change in various exogenous variables.

The regression coefficients for death rates and child school attendance (Education A) do not differ greatly for the two subperiods, but those for adult education (Education C), female activity, and family workers distinctly differ. The sum of the coefficients for the two prior years' death rates increases from 1.3 in the early subperiod to 1.9 in the later.[2] In the long run a decline in death rates appears to be associated with a reduction in the rate of population growth, because the subsequent fall in birth rates more than compensates for the increased size of the surviving population. Reducing death rates by improving health and economic conditions may not have exacerbated but on the contrary may have facilitated the process of demographic transition in Puerto Rico, slackening the high rate of population growth.[3]

The effect of child school attendance on birth rates seems to be relatively constant in the two subperiods, but the implied effect of adult education diminishes markedly in the later subperiod. This result could stem from the initially close association between education and

---

[1]Description of estimation procedure is in Appendix D.

[2]Since the absolute size of the decline in death rates is smaller in the later subperiod, one could infer that the form of the model might better be expressed in terms of relative changes or in logarithmic form. Alternatively, if parents seek a surviving family size, as hypothesized, the inverse of the death rate as an approximation for the survival rate could be incorporated into the linear model, since the survival rate would reasonably enter into the parents' linear adjustment function.

[3]Other researchers have also come to the conclusion that death rates may have to fall to induce a general decline in the rate of population growth. This conclusion develops out of research using computer simulation techniques done by David M. Heer and Dean O. Smith, "Mortality Level, Desired Family Size and Population Increase," revised version of paper presented at meeting of the Population Association of America, Cincinnati, Ohio, April 1967. Also see H. Fredericksen, "Determinants and Consequences of Mortality-and Fertility Trends," Public Health Reports, Vol. 81, pp. 715-727, U.S. Public Health Service, Washington, D.C., 1966.

Table 10

SUBPERIOD REGRESSION RESULTS: BIRTH RATES IN PUERTO RICO, 1950s[a]

(standard errors of regression coefficients shown in parentheses)[b]

Regression Number	Period	Estimation Technique	Constant Term	Lagged Death Rates		Education		Female Activity	Family Worker	ρ	R^2
				t-1	t-2	(A)	(C)				
1	1951-1954	Original	36.20	.960[e] (.202)	.564[d] (.205)	-.127[c] (.050)	-1.935[e] (.408)	-.189[e] (.035)	.757[e] (.117)	.265	.496
2	1951-1954	Transformed	25.96	.869[e] (.201)	.634[d] (.203)	-.113[c] (.048)	-1.897[e] (.388)	-.165[e] (.035)	.764[e] (.112)	n.a.	.492
3	1955-1957	Original	30.45	1.048[e] (.302)	.791[d] (.277)	-.133[d] (.050)	-.820[c] (.393)	-.116[c] (.050)	1.570[e] (.171)	.354	.534
4	1955-1957	Transformed	21.30	.981[d] (.298)	.966[e] (.274)	-.120[c] (.051)	-.797[c] (.378)	-.063 (.054)	1.509[e] (.167)	n.a.	.489
5	1951-1957	Original	36.15	.860[e] (.173)	.324[c] (.164)	-.088[c] (.037)	-1.575[e] (.300)	-.178[e] (.030)	.749[e] (.093)	.126	.460
6	1951-1957	Transformed	26.88	.835[e] (.171)	.304 (.163)	-.083[c] (.037)	-1.451[e] (.296)	-.179[e] (.030)	.766[e] (.095)	n.a.	.449

Notes:

n.a. - Not appropriate.

[a]Definitions of variables are given in Table 5.

[b]The standard errors are asymptotically unbiased for the transformed estimates. Number of observations = 525.

[c]Regression coefficient is significant at .01.

[d]Regression coefficient is significant at .01.

[e]Regression coefficient is significant at .001.

Table 11

ELASTICITY OF MUNICIPAL BIRTH RATES WITH RESPECT TO EXOGENOUS VARIABLES: PUERTO RICO SELECTED PERIODS[a]

Variable[b]	1951-1954	1955-1957	1951-1957
Death Rate (t-1)	.224	.206	.194[c]
Death Rate (t-2)	.178	.213	.077[c]
Summed long-run effect	.402	.419	.272
Education			
(A) School Attendance Rate	-.266	-.284	-.181
(C) Adult Schooling (years)	-.192	-.091	-.154
Female Activity Rate	-.090	-.035[c]	-.098
Family Workers (unpaid)	.084	.109	.072

Notes:

[a]Elasticities represent the relative association between a small change in the exogenous variables and the endogenous variable (birth rates), both measured at the means of the original date.

[b]Definitions of variables in Table 5.

[c]Regression coefficient is not asymptotically significant at .05 level.

knowledge of modern methods of birth control which was probably weakened by the influence of the national family planning program. The decreasing size and significance of the estimated coefficient for female activity rates in the later subperiod may be due to the exodus of young working women to mainland United States in the early 1950s which explains the atypical decline in the proportion of adult women in the labor force by 1960 compared to 1950. Unpaid family workers as a percentage of the civilian labor force also fell in this period from 3.1 to .6 percent. If this variable continued to reflect regional attitudes toward child-labor practices, even though these activities were reimbursed within or outside of the home, the estimated coefficient for this variable would increase as the actual number of unpaid family workers decreased. These and other hypotheses ventured to explain changes in the magnitude and significance of the component variables in the family planning model either involve speculations or require extensive additional research that take us beyond the scope of this Memorandum. It can be inferred, however, from the subperiod estimates that the family planning model in this linear formulation is not invariant over time. Alternative specifications of the model should be considered.

In summary, four environmental variables lagged one year and death rates lagged one and two years constitute the core of the family planning model. Additional environmental and demographic variables, except for personal income, which is marred by a variety of weaknesses, do not account for a significant portion of the remaining unexplained inter-regional variation in birth rates. The estimated effect of each of the central variables is consistent with the family planning hypothesis that assumes parents exercise some control over their births, adjusting their demand for a number of surviving offspring in accordance with economic and demographic opportunities and constraints. The salient features of the family's environment -- mortality or health, education and economic activity of women and children -- account for almost half of the variation in birth rates among 75 municipalities over seven years.

V. SUMMARY AND CONCLUSIONS

THE HYPOTHESIS

The hypothesis was advanced that differences in human environment modify in a systematic manner the number of surviving children parents want, and as parents succeed in at least partially adjusting their reproductive behavior to achieve their family size goal, environmental factors can be interpreted as a cause for observed differences in reproductive behavior. That environment influences human reproductive behavior is self-evident. But that a few quantifiable economic and demographic aspects of the environment add much to our understanding of the determinants of reproductive behavior is controversial. The question raised and examined in this Memorandum is, therefore, one of degree: how helpful are these features of the environment in statistically accounting for observed differences over time and space in reproductive behavior? The many-dimensioned character of the human environment is an essential element of this approach, from which follows the corollary that single-factor analyses are not well designed to probe the complex interplay between environment and reproductive behavior.

The analytical framework developed here focuses on several conceptual variables for which empirical counterparts are not readily available: the death of children; the economic resources of the family that grow out of the efforts of husband, wife, and child, each of whose contribution affects differently the relative attractiveness of children; the educational system that has a direct bearing on the cost of rearing children, influences the achievement goals parents set for their offspring, and may also explain differential exposure to modern methods of birth control; and finally, the web of social and public policies that shape the function of the family and the role of children therein, altering the costs and benefits of rearing a large or small number of children.

THE EVIDENCE

Empirical evidence has been assembled for Puerto Rico to test the principal implications of the family planning hypothesis applied to a

cross section of regions. The hypothesized association between birth rates and death rates in the recent past and selected environmental variables is found to exist in the 1890s when birth rates are seriously underreported, and in the 1950s when demographic and economic materials are more reliable. A generalized least squares estimation procedure proposed by Nerlove is used to reestimate the model taking account of the disturbance effects specific to the regions in the pooled cross-sections. The severity of regional disturbances in the data for the 1890s does not permit one to distinguish with confidence the associations implied by the family planning hypothesis. In the 1950s, however, the generalized least squares estimates are not substantially different from those derived directly by ordinary least squares. It appears that this procedure of Nerlove's may be useful for distinguishing between good and poor data, or between well- and ill-specified models based on pooled cross sections.

The association between birth rates and lagged death rates in the 1950s is distributed over several preceding years, but it is not possible to delineate with confidence the form of this distributed lag from direct multiple regression techniques. Two prior years are included in the final model to take account of the long-run lagged effect of death rates on birth rates. The model implying geometrically declining weights does not appear to fit these data well, though other lag functions deserve further investigation.

School attendance rates among young children and the educational attainment of adults are each associated with lower birth rates, as increased employment of unpaid family workers is related to higher birth rates. Female participation rates, as a proxy for the opportunity value of a woman's time, are also associated with lower birth rates. Finally, median family income was associated with lower birth rates, but the conceptual ambiguity and empirical shortcomings of this variable led to its omission from the final analysis.

Other factors that are often linked to birth rate differentials, such as urban versus rural residence and nonagricultural versus agricultural livelihoods, appear to be unrelated to differences in regional

birth rates in Puerto Rico when the variables in the family planning model are accounted for.

Simple experimentation with demographic variables, such as those based on the age/sex composition of the population and the proportion of adults married, did not establish that these factors were important in predicting interregional variations in crude birth rates in Puerto Rico either in the 1890s or 1950s.

ANALYTICAL SHORTCOMINGS

The empirical variables used in this analysis do not always reflect accurately the features of the family environment that are hypothesized to alter the net costs of children. The model implied that the death of children to still fertile parents might induce within these families, after a brief lag, somewhat higher birth rates. As a proxy for this specific child death rate, the crude death rate was used, even though the association between these two measures of mortality weakens as demographic transition progresses and a growing proportion of all deaths take the lives of the old rather than the young. When the health and living standards of a population improve, and the composition of the population ages, the infant mortality rate may become a better reflection than the crude death rate of the specific child death rate. Further study is needed to appraise the relative merits of these and other empirical approximations for the child death rate in various circumstances.

Refinement of other empirical variables also seems called for, but these adjustments require information that is not now available in the Puerto Rican case, and is wholly lacking in most less developed countries. Income should be measured as real "opportunity income" including all goods and services earned per unit of time spent in economic activity or search of it. Opportunity income per unit time should be evaluated separately for both the husband and wife, for economic logic suggests that the effects of these two sources of family income may be quite different on the decision to have more children. The value of the productive services rendered by children to their parents is important in the context of the family planning model, but this is a

difficult number to derive or estimate from standard statistical sources. It would also be advantageous to study the role of birth control knowledge that reduces the cost and inconvenience associated with limiting unwanted births. The inclusion of this variable would reduce spurious correlations that attribute the effect of birth control knowledge to other factors such as general education and income. These refinements and generalizations of the empirical variables in a family planning model would represent an important further step in evaluating the implications of the hypothesis.

A second area of research that deserves attention is the specification of the statistical model. The conceptual framework implies only the direction of the direct effect on birth rates of various exogenous variables, and intuition and fragmentary data suggest the general timing of their effect. However, little is known of the precise way these exogenous variables interact to alter family planning decisions over time, and how in turn these decisions once made are translated into an observed change in birth rates. The model set forth in Section III is premised on many highly arbitrary assumptions about this process that though reasonable and convenient should be varied. For example, why should all variables enter the birth function linearly and independent of one another; might not interactive terms be more plausible or a multiplicative formulation of the entire model which would lead to a logarithmic estimation equation? The link between the static formulation and the dynamic adjustment mechanism is conceived here in terms of common econometric forms that may not be altogether appropriate for modeling this area of human behavior.

Finally, research must contend with data from a variety of environments, though much of these data will be necessarily inferior to those available for contemporary Puerto Rico. If the conceptual approach embodied in the family planning hypothesis has merit, it is presumably as an analytic tool for the general study of population growth and fertility. Should the implications of this analysis of Puerto Rican data have some general validity to other countries and times, the approach may help to account for change in birth rates over time that

have heretofore been quite unpredictable. Knowledge of the determinants of fertility may ultimately permit society to influence its rate of population growth through indirect policies without sacrificing individual choice in this important area. At this preliminary stage of study, the urgent need is for much experimentation in empirical and statistical work to specify adequately this most intricate area of human behavior in terms that are relatively simple and abstract but useful.

Appendix A

ASPECTS OF PUERTO RICAN CASE STUDY

This Appendix presents some evidence on several implications of this study. First, there is the hypothesis that women's economic opportunities influence their allocation of time, and in particular that their opportunity income affects their choice between earning income and bearing and rearing children. Other factors in addition to the opportunity to earn extra income influence this important choice, namely, their permanent consumption level which is typically related to their husband's income level, and their subjective evaluation of leisure and family activities vis-à-vis income earning activities. Since these factors are all interrelated, one should deal simultaneously with the effect of all to estimate what effect any one may have on labor force participation or reproductive behavior. Cain and Mincer have explored the determinants of participation rates among U.S. women, confirming in cross sections the expected effect of women's wages, unemployment, husband's income, education, and young children in the home.[1] Cain's preliminary investigation of cross-sectional associations between economic variables and completed fertility in metropolitan United States suggests that completed fertility is, as predicted, negatively associated with women's participation rates and wages.[2] The simultaneous

---

[1] Glen G. Cain, Married Women in the Labor Force: An Economic Analysis, University of Chicago Press, Chicago, 1966; and Jacob Mincer, "The Labor Force Participation of Married Women," Aspects of Labor Economics, National Bureau of Economic Research, Princeton University Press, Princeton, 1962.

[2] Glen G. Cain and Adriana Weininger, "Economic Determinants of Fertility. Preliminary Results from Cross-Sections, Aggregative Data," draft paper, April 6, 1967. A variety of other evidence confirms the effect of education on birth intervals and the concentration of child bearing in order to participate. See Pascal Whelpton's "Trends and Differentials in Spacing of Births," Demography, Vol. 1, No. 1, 1964, pp. 83-93, in particular Table 4 on birth intervals by education and cohort. For international comparisons of age-fertility relationship, the statistics can be found in the Population Bulletin of the United Nations, No. 7, 1963, Chapter VII. Examples of countries where the "early-peaked" age fertility pattern is noticeable are Bulgaria, East

consideration of all of these factors in either the United States or Puerto Rico setting is a task for another and larger study. But some aggregate data from these two environments hints at the types of differences one may expect in these relationships between more and less developed countries.

Education appears to enhance the marketable value of a laborer's skills, and thus education and income tend to be associated for both men and women. But the relative scarcity of highly skilled labor and the relative abundance of unskilled labor in the less developed countries implies that if the demand for skilled labor were similar in both types of countries, the differences in the marginal productivity and earnings associated with levels of educational attainment and skill would be substantially greater in the less developed country than they are in the more developed one. The comparison between median income levels of women by educational attainment in Puerto Rico and the United States is consistent with this inference, as seen in columns 6 and 7 of Table A-1. To confirm this inference rigorously, earnings per unit time in the labor force would be required, but the data for this test are not available from published sources.

In San Juan, a high school graduate has a median income more than nine times that received by a woman with no schooling; in the United States the ratio is three to one. The relative income differences associated with education level are much greater for women in San Juan than for those in urban United States, and except for college graduates even the absolute increases in median income associated with additional years of schooling are greater for women in San Juan than for those in the United States. Other data can be cited from cities in less developed countries that suggest wide income-education differentials are a general pattern in less developed countries.[1]

Germany, Hungary, Poland, United States, Finland, Norway, Sweden, Denmark, and France. Most of these countries are also noteworthy for high female participation rates. See Some Factors in Economic Growth in Europe During the 1950s, United Nations Economic Commission for Europe, Geneva, 1964, Chap. IV, Table 3, p. 5.

[1]From a sample survey of Bogota, taken in 1964, a similar pattern emerges where women with 11 years of schooling earn incomes more than

Table A-1

LABOR FORCE PARTICIPATION RATES, NUMBER OF CHILDREN EVER BORN, AND MEDIAN INCOME OF WOMEN IN 1960: SAN JUAN, PUERTO RICO, AND THE UNITED STATES

Educational Attainment	Married Women Age 35-44 Participation Rates		Number of Children Ever Born to Women Ever Married, Age 35 yrs.			Median Income of Women 25 Years of Age or Older in Urban Areas	
	San Juan SMSA[a]	U.S.-SMSA's Central Cities[a]	Puerto Rico	San Juan	U.S.-Urban	Puerto Rico	United States
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
No schooling	11.7	29.1	7.04	5.83		180	731
1-2 years elementary	16.3	37.2	6.83	5.55	2.44	287	814
3-4 years elementary			6.55			344	
5-7 years elementary	21.6[b]	37.6	5.33	4.49[b]		540	997
8 years elementary	23.5[b]	38.7	3.93	3.49[b]	2.47	846	1299
1-3 years high school	31.8	39.1	3.37	2.96	2.45	1176	1764
4 years high school	38.6	39.2	2.66	2.36	2.23	1726	2363
1-3 years college	53.0	39.1	2.63	2.34	2.24	2026	2571
4 years college		38.5	2.58		2.19	2763	3873
5 or more college		58.2					
Total	26.9	39.2	5.58	4.06	--	462	1730

Notes:

-- Not available.

[a]SMSA: Standard Metropolitan Statistical Area in U.S. Census of Population.

[b]Breakdown 5-6 and 7-8 years of schooling for these entries.

Sources:

Columns 1, 2, and 4 are from special tabulations presented by Robert O. Carleton, "Labor Force Participation: A Stimulus to Fertility in Puerto Rico," Demography, Vol. 2, 1965, pp. 233-239, Tables 1, 2, and 4.

Columns 3 and 6 are from U.S. Census of Population: 1960, Vol. I, Characteristics of the Population, Part 53, Tables 95 and 117, pp. 53-276 and 387.

Column 5 from U.S. Census of Population: 1960, Subject Reports: Women by Number of Children Ever Born, PC(2)-3A, Table 28, and Table 29, pp. 119-132.

Column 7 from U.S. Census of Population: 1960, Vol. I, Characteristics of the Population, Part 1, Summary United States, Table 223, pp. 1-590.

Participation rates in the labor force for women and completed fertility are also related to educational attainment and income, though the direction of causation cannot be established firmly. Table A-1 shows that participation rates are higher and fertility lower the more education married women age 35 to 44 have. In all probability, the greater differences in participation rates and fertility associated with education level in Puerto Rico than in the United States stem in some part from the much larger income differences associated with education level.

But the picture is more complex than this two-way breakdown portrays. Better educated women with some income may work more each year than the less educated, and thus the income differences shown in Table 9 may overstate wage (per unit time) differences. Other socio-economic links may be important. Educated women tend to marry husbands who earn higher incomes, reducing their incentive to earn extra income themselves. Also, education is associated with higher divorce and separation rates, which might have the opposite effect on participation rates. With regard to fertility rates, early child death is less frequent among better educated high socio-economic classes, and consequently the differences in number of children ever born, shown in Table A-1, overstate the differences in number of surviving children, our proxy for family size goal.

A second issue raised in this study is the probable impact of changes in the relative earning capacity of men and women on the organization of the family unit, and on the number of children parents want and can afford. It was argued that if women's opportunity to earn income increased relative to men's this could have the consequence of increasing female participation in the labor force, adding to their economic independence, and reducing their traditional desires for children and a conventionally formalized union. The resulting more fluid family arrangements would reinforce demands for lower birth rates.

---

three times greater than those with no schooling, and a college graduate earned five times what a woman with five years of schooling earned. These Bogota data referred to earnings per month for those in the labor force, and are consequently more likely to represent conceptually opportunity income per unit time in the labor force.

To compare the opportunity income of men and women, earnings data are required for both sexes for comparable time worked and for comparable occupational skills. This first condition is important since women tend more than men to work part of the year, and consequently their reported annual earnings or income understate, relative to men, their earning capacity. Table A-2 presents the ratio of female to male median income by weeks worked in 1959 for the United States and Puerto Rico. In the United States women receive about 60 percent as much as men for a comparable number of weeks worked; but for all persons with income regardless of work history, women report a median income only one-third that of men. In Puerto Rico, the incomes of men and women are approximately the same when they work similar numbers of weeks, but over all, women report less than half the median income of men. The near equality of income between the sexes for comparable time worked in Puerto Rico nevertheless neglects skill and age differences between sexes. It appears that the educational attainment of the female labor force is on average greater than that of the male labor force, and thus women in Puerto Rico may still be paid somewhat less than men for "equal" work, though clearly the gap is far narrower than in the United States. Data for the labor force in Bogota, Colombia shows the same pattern, with women receiving approximately the same monthly earnings as men with similar educational attainment. Though these two fragments of data are only tentative, they do suggest that the disparity in remuneration of the sexes that occurs in developed countries may not be as great in less developed countries, at least once the traditional segregation of the labor market for men and women has eroded under the pressures of urbanization and industrialization. Further study is needed to determine if this shift in relative economic status of the sexes accounts for differences in family structure and fertility as implied by the analytical framework here.

A final subject deserves attention, that of the effect of external migration on Puerto Rican birth rates. In some discussions of Puerto Rican fertility, it is alleged that the postwar decline in crude birth rate is largely a function of the migration of youth to mainland United

Table A-2

RATIO OF FEMALE TO MALE MEDIAN INCOMES BY AGE AND WEEKS WORKED IN 1959: PUERTO RICO AND UNITED STATES

Persons, by Age and Weeks Worked	Puerto Rico	United States
All persons 25 years of age or over with income	.43	.33
All persons 14 years of age or over	.48	.35
All persons 14 years of age or over and working in 1959	.87	.43
Working 50-52 weeks	.86	.59
Working 48-49 weeks	1.01	.56
Working 40-47 weeks	1.07	
Working 27-39 weeks	1.05	.61
Working 14-26 weeks	.89	.62
Working 13 weeks or less	.69	.80

Source:

U.S. Census of Population: 1960, Vol. I, Characteristics of the Population, Part 53, Puerto Rico, Tables 115 and 117, pp. 53-384 and 387; part 1, United States Summary, Tables 221 and 223, pp. 1-587 and 590.

States. This migration flow reached major proportions during the mid-1950s. To estimate the effect of migration on the crude birth rate, information is first needed on the net number of migrants leaving Puerto Rico by age and sex, and then one must assume that those leaving Puerto Rico would have sustained birth rates (and death rates) similar to the population that remained on the island. This approach is defensible, though the assumption on how migrants behave probably overstates the effect of migration on birth rates, since in other historical instances migrants have tended to have lower than average birth rates. Regardless, the age/sex data on migrants appears to be lacking, so this exercise cannot be performed.

Another question, however, can be answered, though it does not directly measure the marginal effect of migration on crude birth rates. Has a substantial fraction of the recent decline in crude birth rates in Puerto Rico been due to changes in age structure that were linked to external migration? Using the age structure of the population reported in the 1950 census, age-specific birth rates for the island can be applied to this constant age structure to derive an age-standardized birth rate. This age-standardized birth rate along with the crude birth rate and the age-specific birth rates are shown in Table A-3 for selected years. Births in 1939-1941 are likely to be somewhat underreported, but no evidence could be found on the magnitude of this error.

These various measures of birth rates in Puerto Rico indicate that the changes in age structure have had a noticeable effect on the crude birth rate before 1950 and since 1958. In the period of heaviest migration, however, in the early and mid-1950s, the crude and standardized birth rates move together. Although it cannot be stated with assurance that the impact of migration on the age structure and consequently on the crude birth rates was negligible, the opposite contention requires further support.

Table A-3

AGE-SPECIFIC, CRUDE, AND STANDARDIZED BIRTH RATES PER THOUSAND IN PUERTO RICO FOR SELECTED YEARS

Age Group (Years)	1939-1941	1950	1954	1958	1963
Less than 15[a]	.2	.6	.6	.6	.3
15-19	77.9	99.2	97.5	104.3	101.4
20-24	247.8	279.7	254.6	291.1	257.9
25-29	261.9	260.3	218.9	202.7	191.3
30-34	215.1	200.0	159.9	133.1	120.8
35-39	140.1	143.1	131.3	103.9	94.9
40-44	43.3	53.1	47.5	41.5	9.9
45 or more[b]	9.1	11.7	10.8	9.7[c]	
Crude birth rate	39.2	39.0	35.2	33.2	30.7
Standardized birth rate, 1950 age composition	36.6	38.9	34.5	34.2	34.0

Notes:

[a] Births divided by population 10 to 14.

[b] Births divided by population 45 to 49.

[c] Births only for ages 45-49.

Sources:

UN Demographic Yearbook for 1939/41 and 1950; Vital Statistics of the United States, Vol. I, Natality, 1954, 1958, 1963.

Appendix B

THE TREATMENT OF UNCERTAINTY IN THE FAMILY FORMATION PROCESS

The view of uncertainty adopted here corresponds to the expected utility theory of decisionmaking under uncertainty as developed extensively in the literature of economics and statistics.[1] A conditional probability density function represents the chance that various numbers of children will survive given any planned birth schedule (contraceptive practices) and mortality regime.[2] To assign a single value to this set of expected outcomes to compare it with alternative outcomes that result from other planned birth schedules, parents are assumed to weigh the subjective value of each outcome (from their net-cost schedule for children) by the probability of its occurring given the goal sought. This yields a measure of the expected value of the probable outcomes. Finally, it is assumed that parents gravitate toward the birth schedule that yields them the lowest expected cost or highest expected benefit.

Given reasonable assumptions on the incidence of death and the number of births parents can expect to have given their family size goal and mortality regime, the conditional probability density function of surviving children appears to be approximately symmetric with

---

[1] A general description of the expected utility theory can be found in a number of textbooks: Robert Schlaifer, Introduction to Statistics for Business Decisions, McGraw Hill, New York, 1961; R. Duncan Luce and Howard Raiffa, Games and Decisions: Introduction and Critical Survey, John Wiley and Sons, New York, 1957, Ch. 2.

[2] For simplicity it is assumed here that family planning decisions are not made sequentially. In fact, family goals and reproductive behavior are probably revised many times in the course of family formation. It is probable that the loss of a child leads to a reevaluation of the desirability of another birth, implying that the initial expectations held by the parent about the child's chance for survival are constantly being revised.

It is also assumed that on balance boys and girls are equally valued by parents, regardless of the prior sex distribution of children. Although this assumption is probably not realistic, the explicit treatment of the sex distribution of children does not appear to make a significant change in the analytic approach followed here.

respect to the number of children wanted.[1] Were this the general case, the effect of uncertainty on parents' decision of how many births to seek would be a function of the asymmetry of their net-cost schedule with respect to their family size goal. For example, when parents emphasize having at least a certain number of surviving children and do not regard too many as much of a liability, uncertainty would induce parents to raise their desired birth rate, and the reverse effect on desired birth rates would be expected if parents regarded too many children as a greater hardship than too few.

The effect of changing the level of mortality or the effectiveness of contraceptive practices would be to change the variance of the conditional probability density function. The greater variance of the probability function that would be associated with higher death rates and poorer contraception practices implies greater uncertainty in the family planning process, and a larger effect on the decisionmaking

---

[1] The survival probability of each child is assumed to be dependent only on the child death rate, and to be statistically independent of his siblings' fate, so that the expectations for survival from any given number of births can be calculated by using the binomial distribution (three place tables are presented in F. Mosteller, R. E. K. Rourke, and G. B. Thomas, Jr., Probability with Statistical Applications, Addison-Wesley Publishing Company, Reading, Massachusetts, 1961, Tables IV-A, pp. 433-443). Two survival probabilities for children from birth to age 15 were investigated, .7 and .9. No empirical evidence directly pertains to the other source of uncertainty in the family planning process, the birth rate. From several sources on fertility and sterility among American women, two estimates were constructed for the number of births a couple could expect given a goal of four surviving children and modern or traditional practices of contraception. The results are only rough approximations given the lack of empirical materials on this subject:

Form of	Percent chance that the number of births will be exactly									
Contraception	0	1	2	3	4	5	6	7	8	9 or more
Traditional	1.5	2.0	8.1	21.0	25.0	21.5	11.0	5.5	3.0	1.5
Modern	1.0	1.5	8.0	20.0	55.0	7.0	4.0	2.0	.7	.3

Combining the traditional outcome given above with the .7 probability of child survival, and the modern outcome with the .9 survival factor, two conditional probability density functions were derived and both had approximately symmetric properties with respect to the goal of four surviving children. The variance of the probability function was much greater for the high death, traditional contraception case than for the low death, modern contraception case.

process if parents' net cost schedules for children are asymmetric about their family size goal.

For an empirical measure of uncertainty, historical variation in local death rates are considered for the prior seven years. A linear trend in time is included in a least squares regression on local death rates properly scaled for the size of the local population. The unexplained variation in death rates over these seven years is used as a proxy for uncertainty in the family formation process that stems from the incidence of death in the region.[1] This estimate is used in the regression results reported in Table 8.

---

[1]The estimates were computed for the previous seven years for the equation:

$$\frac{D_{it}}{\sqrt{n_{it}}} = \alpha + \beta \frac{t}{\sqrt{n_{it}}} + e$$

where D_{it} is the death rate in the i^{th} region in time period t, t=1, 2,...,7, n_{it} is the population in the i^{th} region in time t, e is a disturbance term and the parameters α and β are estimated by least squares. The measure of uncertainty used in this Memorandum is the estimated variance of "e" from the above regressions for each year and region, or in other words, the variance of the residuals.

77

APPENDIX C

Table C-1

VARIABLES IN EARLY REGRESSION: 1894-1897[a]

Variable	Name	Mean	Std Deviation	Source of Data
1	Birth rates (per thousand inhabitants)	28.01	11.23	Report on the Census of Puerto Rico, 1899, War Dept., Office Director Census of Puerto Rico, Washington, D.C., 1900, Table XXXV, pp. 345-347.
2	Death rates[b] (per thousand inhabitants)	26.24	7.20	Table XXXVII, pp. 350-351.
3	Education variables (percent)			
	A. School attendance 7-9	7.04	4.63	Table XVII, pp. 243-245.
	B. School attendance 10-17	8.46	3.72	
	C. Literacy persons 10 or more	15.56	6.44	
4	Female activity (percent)	18.20	10.70	Table XXIII, pp. 382-305.
5	Urbanization (percent)	17.67	15.67	Table IV, p. 163.
6	Married proportion (percent)	51.73	3.35	Table XIII, pp. 194-195.
7	Age-sex compositional proxy (percent)	19.55	1.23	Table VIII, pp. 172-173.

Notes:

[a]Variables are defined in Table 1.

[b]Death rates lagged one year.

Table C-2

VARIABLES IN MODERN REGRESSIONS: 1951-1957[a]

Variable	Name	Mean	Std Deviation	Source of Data
1	Birth rates (per thousand inhabitants)	32.83	6.14	Various numbers of the Anuario Estadistico Puerto Rico, Puerto Rico Planning Board, Bureau of Economics and Statistics, San Juan, and unpublished data provided by the Dept. of Public Health and Vital Registry, San Juan, Puerto Rico.
2	Death rates[b] (per thousand inhabitants)	7.83	2.00	Same as birth rates.
3	Education variables (percent)			
	A. School attendance 7-13[b]	71.95	6.28	U.S. Census of Population: 1950, Vol. II, Characteristics of the Population, Part 51-54, Table 38, pp. 53-72; U.S. Census of Population, Part 53, Table 68, pp. 53-179.
	B. School attendance 14-17	47.67	7.92	Same as above.
	C. Adult schooling[b] (median years)	3.50	.77	Census, 1950, Table 10, pp. 53-25; Census, 1960, Table 35, pp. 53-116.
4	Income (current dollars)[b]	$386.	145.	Census, 1950, Table 39, pp. 53-78; Census 1960, Table 71, pp. 53-206.
5	Uncertainty	.268	.245	Unexplained variance in prior death rates.[c]
6	Unpaid family workers[b] (percent)	3.10	2.36	Census 1950, Table 40, pp. 53-84; Census 1960, Table 69, pp. 53-188.

Table C-2 (continued)

Variable	Name	Mean	Std Deviation	Source of Data
7	Female activity[b] (percent)	18.09	7.31	Census 1950, Table 10, pp. 53-25; Census 1960, Table 68, pp. 53-186.
8	Agriculture[b] (percent)	20.49	16.94	Census 1950, Table 40, pp. 53-84; Census 1960, Table 69, pp. 53-188.
9	Urbanization[b] (percent)	20.49	21.54	Census 1950, Table 10, pp. 53-25; Census 1960, Table 5, pp. 53-12.
10	Married Proportion[b] (percent)	67.33	2.96	Census 1950, Table 34, pp. 53- ; Census 1960, Table 29, pp. 53-93.
11	Age-sex compositional proxy			
	1950	36.23	2.60	Census 1950, Table 34, pp. 53-49; Census 1960, Table 28, pp. 53-74; Demographic Yearbook 1955, United Nations, Table 21, p. 628 and Vital Statistics of the United States 1960, Vol. I, Natality, Table 4-8, p. 4-5.[d]
	1960	30.45	2.62	

Notes:

[a]Variables defined in Table 5 and notes below.

[b]Mean and standard deviations of these variables pertain to the series lagged one year.

[c]This variable is the residual variance of death rates in the prior seven years (for example, for 1951, death rates are used for 1944 to 1950) not accounted for a linear time trend. Least squares estimates of the residual variance are used here.

[d]This variable is constructed by weighting the six nationwide female-age-specific birth rates by the number of thousand women in each age-specific cohort in the municipality. If the nationwide age-specific birth rates were applicable to each municipio, this proxy would conform to the observed birth rate in each municipio.

Appendix D

ESTIMATION PROCEDURE FOR DYNAMIC RELATIONSHIP BASED ON A COMBINATION OF TIME SERIES AND CROSS SECTION DATA[1]

Following Balestra and Nerlove, consider the relationship for pooled cross-sectional and time series data

$$y = X\beta + u \tag{D.1}$$

where y is a NT x 1 vector of endogenous variables, X is a NT x K matrix of exogenous variables, and u is a NT x 1 vector of disturbances. Also, β is a K x 1 vector of constant coefficients of the exogenous variables. Observations of this relationship are drawn from N cross-sectional "regions" for T time periods. It is assumed that the disturbance u_{it} for the i^{th} region, i=1,...,N, and t^{th} time period, t=1,...,T, may be decomposed into two statistically independent parts: a regional effect and a remainder

$$u_{it} = \mu_i + \nu_{it} \tag{D.2}$$

The random variables μ_i and ν_{it} are assumed to have zero means and to be independent.

$$E\,\mu_i\nu_{it} = 0, \text{ for all } i,t \tag{D.3}$$

and no serial correlation among ν_{it}, and that these are independent among regions

$$E\,\nu_{it}\nu_{i't'} = \begin{cases} \sigma_\nu^2, & i=i' \text{ and } t=t' \\ 0, & \text{otherwise.} \end{cases} \tag{D.4}$$

---

[1] Most of this section is derived directly from the text and appendix of Pietro Balestra and Marc Nerlove, "Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas," Econometrica, Vol. 34, No. 3, July 1966, and the remainder proposed in discussions with Nerlove.

similarly

$$E\,\mu_i\mu_{i'} = \begin{cases} \sigma_\mu^2, & i=i' \\ 0, & \text{otherwise.} \end{cases} \tag{D.5}$$

Equations (D.3) imply that the variance covariance matrix of disturbance u may be written in block diagonal form

$$E\,uu' = \Omega = \sigma^2 \begin{bmatrix} A & 0 & \dots & 0 \\ 0 & A & \dots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \dots & A \end{bmatrix} \tag{D.6}$$

where Ω is a NT x NT matrix, and

$$Eu_i u_{i'} = \sigma^2 A = \sigma^2 \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \rho & \rho & \dots & 1 \end{bmatrix} \tag{D.7}$$

where A is a T x T matrix, and

$$\begin{aligned} \sigma^2 &= \sigma_\mu^2 + \sigma_\nu^2 \\ \rho &= \sigma_\mu^2 \,/\, \sigma^2 . \end{aligned} \tag{D.8}$$

Where no lagged values of the endogenous variables are included among the explanatory variables, the least squares estimates of β in (D.1) are consistent and asymptotically unbiased. But these estimates are nevertheless not minimum variance or, in general, asymptotically efficient. The object of this appendix is to derive the transformation for (D.1)

$$Zy = ZX\beta + Zu \tag{D.9}$$

where Z is an orthogonal transformation matrix, NT by NT, subject to the condition that the transformed disturbances possess a variance-covariance matrix appropriate for least squares estimation,

$$E\ Zuu'Z' = \sigma^2\ I \tag{D.10}$$

where I is the identity matrix. Then the ordinary least squares estimates of β^* in (D.9) are consistent, and both asymptotically unbiased and efficient.[1]

$$\beta^* = [(ZX)'ZX]^{-1}\ (ZX)'ZY \tag{D.11}$$

or

$$\beta^* = (X'Z'ZX)^{-1}\ X'Z'ZY\ . \tag{D.12}$$

Let C be a T x T orthogonal matrix, every element of the first row of which is a $1/\sqrt{T}$. If e is a t x 1 vector consisting of ones, C may be written as

$$C = \begin{bmatrix} e'/\sqrt{T} \\ C_1 \end{bmatrix} \tag{D.13}$$

imposing the conditions on C_1 that

$$\begin{aligned} C_1 e &= 0 \\ C_1 C_1' &= I \\ C_1' C_1 &= I - ee'/T. \end{aligned} \tag{D.14}$$

There is an infinite number of such matrices C_1 that satisfy the above conditions. Balestra and Nerlove give an example.[2]

From (D.3) it follows that

$$A = (1-\rho)I + \rho ee' \tag{D.15}$$

and since C is orthogonal

---

[1] J. Johnston, Econometric Methods, McGraw Hill Book Co., New York, 1963, pp. 186-187.

[2] See their Appendix, p. 609.

$$CAC' = \begin{bmatrix} \xi & 0 & \dots & 0 \\ 0 & \eta & \dots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \dots & \eta \end{bmatrix} \tag{D.16}$$

where

$$\begin{aligned} \xi &= (1-\rho) + T\rho \\ \eta &= (1-\rho) \end{aligned} \tag{D.17}$$

The square root of A, which is positive semi-definite, is defined as that matrix with characteristic roots equal to the square roots of the characteristic roots of A.[1] Thus we define

$$F = CA^{-\frac{1}{2}}C' = \begin{bmatrix} 1/\sqrt{\xi} & 0 & \dots & 0 \\ 0 & 1/\sqrt{\eta} & \dots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & & 1/\sqrt{\eta} \end{bmatrix} \tag{D.18}$$

which may be rewritten

$$F = \left(\frac{1}{\sqrt{\xi}} - \frac{1}{\sqrt{\eta}}\right) \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \dots & 0 \end{bmatrix} + \frac{1}{\sqrt{\eta}} I \tag{D.19}$$

simplifying notation

$$F = \varphi \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \dots & 0 \end{bmatrix} + \Psi I \tag{D.20}$$

from (D.18) and (D.13)

$$A^{-\frac{1}{2}} = C'FC \tag{D.21}$$

---

[1] Richard Bellman, Introduction to Matrix Analysis, McGraw Hill Book Co., New York, 1960, pp. 92-93.

or

$$A^{-\frac{1}{2}} = \left[e'/\sqrt{T} C_1\right] \left[\varphi \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \dots & 0 \end{bmatrix} + \Psi \ I \right] \begin{bmatrix} e'/\sqrt{T} \\ C_1 \end{bmatrix} \tag{D.22}$$

and multiplying through

$$A^{-\frac{1}{2}} = \frac{e}{\sqrt{T}} \varphi \frac{e'}{\sqrt{T}} + \Psi \ I \left\{ \frac{ee'}{T} + C_1'C_1 \right\} \tag{D.23}$$

and using the third condition on C_1 (D.14) it follows that

$$A^{-\frac{1}{2}} = \varphi \frac{ee'}{T} + \Psi \ I \ . \tag{D.24}$$

This represents the orthogonal transformation required for equation (D.1) and can be expressed directly in terms of ξ and η which are determined in (D.17) by ρ and T. The transformed vector of observations for the i^{th} region is thus

$$X_i^* = X_i A^{-\frac{1}{2}} = X_i \left[\varphi \frac{ee'}{T} + \Psi \ I \right] , \tag{D.25}$$

and since

$$\frac{e'X_i}{T} = \bar{X}_i$$

where $\bar{X}_i$ is the mean of Xs from the T observations on the i^{th} region

$$X_i^* = \varphi \ e\bar{X}_i + \Psi \ X_i \ , \tag{D.26}$$

or in terms of ξ and η the T x 1 vector of observations for the i^{th} region for any variable is as follows

$$X_i^* = \begin{bmatrix} \frac{X_{i1} - \bar{X}_i}{\sqrt{\eta}} + \frac{\bar{X}_i}{\sqrt{\xi}} \\ \vdots \\ \frac{X_{it} - \bar{X}_i}{\sqrt{\eta}} + \frac{\bar{X}_i}{\sqrt{\xi}} \end{bmatrix} . \tag{D.27}$$

This transformation, which may be interpreted simply as a weighted average of each observation and its deviation from the mean of that region's observations, depends on ρ, and consistent estimates of β. Ordinary least squares without a lagged endogenous variable yields consistent estimates of β, and from the estimated residuals one may obtain estimates of ρ and σ^2. The maximum likelihood estimate of σ^2 is used here

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{u}_{it}^2}{NT} \qquad (D.28)$$

where $\hat{u}_{it}$ is the estimated residual calculated by least squares from equation (5) of the text.

Three methods for the estimation of ρ, proposed by Balestra and Nerlove, are investigated. The maximum likelihood estimate of ρ has the undesirable property that it may become negative when the variance of regional effects is small in comparison to the overall variance of the residuals. Since a negative ratio of variances, ρ, does not submit to an obvious interpretation, this method is used only as a check on the approximate magnitude of the other two estimates. Dummy variables may also be added to the regression, and the variance of the estimated coefficients for these dummy variables yields an estimate of σ_μ^2, which is, however, biased upward. These two estimates provide a lower and upper bound for the final estimates of ρ computed from a non-negative formula derived by Balestra and Nerlove (in their equation (38)).

$$\rho = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{t'=1}^{T} \hat{u}_{it}\hat{u}_{it'} - \frac{1}{N}\left(\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{u}_{it}\right)^2}{NT\hat{\sigma}^2}. \qquad (D.29)$$

The values of ρ reported in the text, Tables 3 and 8 above, are computed by this formula.[1]

[1]The range of the estimates from these three procedures is not trivial, but some experimentation did not reveal that modest changes of this order in the value of ρ substantially altered the estimates based on the transformed data. For example, in the 1890s, residuals from regression 8 in Table 2 yielded a maximum likelihood estimate of .810, a dummy variable estimate of .905 and a non-negative "adjusted" estimate of .857.