Population Studies, **46** (1992), 103–120
Printed in Great Britain

Estimation of Historical Migration Rates from a Single Census: Interregional Migration in Brazil 1900–1980*

CARL P. SCHMERTMANN†

I. INTRODUCTION

Standard methods for the indirect measurement of historical migration rely on comparisons of populations at several points in time. Can one learn about a country's internal migration history using only a single census?

In this paper I propose a method of analysing common census data for this purpose. Place of birth/place of residence data, when disaggregated by age, contain information on cohorts' migration histories. The single-census method attempts to extract this information by comparing observed lifetime migration rates for each (cohort, origin, destination) combination with those which would obtain under various time trends in period migration rates.

Although it is an imperfect substitute for traditional multiple-census methods, the new approach offers several advantages: (1) it requires fewer data, a fact which may be especially relevant for countries with few reliable censuses; (2) it provides estimates of historical trends in gross, as well as net, rates; and (3) it allows simultaneous analysis of flows between all pairs of regions, rather than requiring that each region be analysed separately in a region against rest-of-country framework.

In order to evaluate the method, I apply it to a country and period for which multiple-census studies exist: twentieth-century Brazil. In tests in which interregional migration data from Brazil's 1980 demographic census are used, the single-census method performs well. The period-effects model fits observed place of birth/place of residence data closely, 'discovers' many of the major trends in twentieth-century Brazilian migration found in earlier studies, and yields several new insights into historical migration patterns. Rough estimates of intercensal net flows for periods up to 80 years before the census date also match fairly closely with those found by researchers who have used standard multiple-census techniques.

II. NOTATION AND DEFINITIONS

Periods and cohorts

From $t = \tau$, the time of the survey, divide the past into a series of n-year periods, denoted by the variable p and numbered according to ending dates. Thus $p = \tau$ represents the n-year period $(\tau-n, \tau)$, $p = \tau-n$ represents the n-year period $(\tau-2n, \tau-n)$, and so on. Throughout the paper, omission of the argument p indicates that a variable refers to the period immediately preceding the survey ($p = \tau$), and omission of t indicates that a variable refers to the survey date ($t = \tau$).

* This paper began as part of my Ph.D. dissertation at Berkeley, and was completed at CEDEPLAR/UFMG in Belo Horizonte, Brazil. I thank Ron Lee for helpful comments, Claudio Machado for generously providing data, and the Rockefeller Foundation for financial support during the paper's completion.

† Center for the Study of Population, Florida State University, Tallahassee, FL 32306-4063, USA.

In each period, divide the population surviving to the end of the period into C n-year age groups, denoted by the variable x and numbered $0, n, 2n, \ldots (C-1)n$ according to the youngest age in the cohort at end of period. Thus $x = 0$ refers to survivors aged $(0, n)$ at end of period, $x = n$ refers to survivors aged $(n, 2n)$, and so on.

Migration hazards and probabilities

Number regions from 1 to R. Assume that during period p, all members of a cohort who survive the period are subject to an identical, constant risk of migrating from s, a sending region, to r, a receiving region. Let $m_x(r, s, p)$ denote the (annual basis) hazard rate during period p for moves from s to r by survivors in age group x. Define matrices $\mathbf{M}_x(p)$ with elements in the rth row and sth column equal to $m_x(r, s, p)$, remembering that columns in these matrices must sum to zero.[1]

Let $\mathbf{P}_x(p)$ denote the transition probability matrix for cohort x during the discrete interval corresponding to period p. The relationship between instantaneous hazards $\mathbf{M}_x(p)$ and discrete transition probabilities $\mathbf{P}_x(p)$ is:[2]

$$\mathbf{P}_x(p) = \exp[n\mathbf{M}_x(p)], \tag{1}$$

where $\exp(\cdot)$ is a matrix operator defined as[3]

$$\exp(\mathbf{A}) = \sum_{k=0}^{\infty} \frac{1}{k!} \mathbf{A}^k. \tag{2}$$

Lifetime migration fractions

Let $f_x(r, s, t)$ denote the lifetime migration fraction from s to r by survivors in age group x at time t – i.e. among those in age group x who were born in s, $f_x(r, s, t)$ is the proportion residing in r at time t. Arrange f values into matrices $\mathbf{F}_x(t)$ as above, noting that columns must sum to unity. Given the above definitions, the expected value of $\mathbf{F}_x(t)$ is

$$\left.\begin{aligned} \mathbf{F}_x(t) &= \mathbf{P}_x(t)\,\mathbf{F}_{x-n}(t-n) \\ &= \mathbf{P}_x(t)\,\mathbf{P}_{x-n}(t-n)\,\mathbf{F}_{x-2n}(t-2n) \\ &= \ldots \\ &= \mathbf{P}_x(t)\,\mathbf{P}_{x-n}(t-n)\ldots\mathbf{P}_0(t-x) \end{aligned}\right\} \tag{3}$$

[1] This restriction implies that one can always recover the elements on the main diagonal of $\mathbf{M}_x$ via

$$m_x(r, r, p) = -\sum_{r \neq s} m_x(r, s, p).$$

In much of the discussion below, I define diagonal terms implicitly in this manner.

[2] For the cohort born during period p (i.e. $x = 0$), the relationship is different, since on average cohort numbers are at risk of migration for only half the period. By analogy with the single-region model, this would yield

$$\mathbf{P}_0(p) = \exp[\tfrac{1}{2}n\mathbf{M}_0(p)].$$

This expression is not strictly correct, however, since simple averaging of hazards over groups is inappropriate when hazards are in matrix form (see N. Keyfitz, *Applied Mathematical Demography* (New York, 1985), p. 356). Nevertheless, when the hazards are all very close to zero, as is the case with interregional migration by 0 to 4-year-olds, the above expression is a very good approximation, and I employ it throughout the paper.

[3] See Keyfitz *op cit.* in fn. 2, or B. Singer and S. Spilerman. 'The representation of social processes by Markov models'. *American Journal of Sociology*, **82** (1976), pp. 1–54 for a more complete discussion of hazard matrices and the associated exp and ln operators.

III. METHOD

Data requirements and assumptions

The method requires an individual-level survey, taken at time τ, which includes information on age (in n-year groups), region of birth, region of residence, and region of residence at time $\tau - n$.[4] It must be possible, through the use of published tables or through direct access to individual records, to generate two sets of data: the C cohort-specific hazard matrices $\mathbf{M}_x$ for the period just before the survey, and the C cohort-specific lifetime fraction matrices $\mathbf{F}_x$ at the survey date.

As implied by the notation above, it must be assumed that migration is a Markov process, that at any given time all individuals in a cohort face an identical set of migration hazards, and that these hazards are constant for each cohort within each n-year period.

Because data exist only for those alive at the survey data τ, all fractions, risks, and trends in the model are necessarily calculated conditional on survival to τ. In order to interpret results as applying to entire past populations, it must be assumed that the probability of survival to τ for all members of a birth cohort is independent of migration history. This requires that the risks of migration and mortality be independent, and that there be no mortality differences across regions. These assumptions are, of course, demonstrably false. However, violations will have relatively small effects on lifetime migration fractions compared to the effects of the central phenomenon under study – namely, variation in migration rates over time.

Logic

Start with a counterfactual question. What if age-specific migration rates did not change over the lifetimes of the cohorts under study? In this case transition probabilities for any age group x in any period p would be identical to those observed for x in the period ending on the census date:[5]

$$\mathbf{P}_x(p) = \mathbf{P}_x \tag{4}$$

and it follows, from (3), that

$$\mathbf{F}_x = \mathbf{P}_x \mathbf{P}_{x-n} \mathbf{P}_{x-2n} \cdots \mathbf{P}_0. \tag{5}$$

In other words, if rates had remained constant, then a cohort's lifetime fractions would simply be the accumulation of current period transition probabilities for itself and for all younger cohorts. This is analogous to the relationship between period and cohort life tables in a stable population.

In practice, migration rates are rarely stable, and the right-hand side of (5) may yield very poor approximations to observed lifetime fractions. For example, in the two graphs in Figure 1 1980 lifetime fractions for flows between two different pairs of Brazilian regions are plotted (solid lines) against the accumulated 1975–80 rates for the same flows (dotted lines). In both cases the discrepancies are large, indicating that some of the assumptions built into (5) must be false. More specifically, the upper graph, in which observed lifetime fractions exceed accumulated current rates for almost all age groups,

[4] Throughout this paper, I define an individual's possible 'states' as geographical regions, and assume that the researcher can observe states at birth, $\tau - n$, and τ. However, the discussion applies to any switching-state model in which states can be observed at three distinct points in time.

[5] The relationship between age- and cohort-specific period rates is considerably more complicated when hazards are in matrix form than it is in the usual single-region model. The logical inference above is straightforward, however: whatever the relationship between age- and cohort-specific rates, if the former are constant over periods then the latter must also be constant.

suggests that 1975–80 migration rates are low by historical standards for this first origin-destination pair. The lower graph suggests the opposite for the second flow: 1975–80 rates are probably much higher than past rates.

Reporting errors, sampling variability, population heterogeneity, and differential mortality clearly contribute to the gaps observed in Figure 1, but given their magnitude it is much likely that the principal explanation lies in changing rates $m_x(r,s,p)$ across periods. Rate changes may be due to a changing age structure of migration rates, to changes in the overall level of migration, or to a combination of these two effects.[6]

Empirical evidence indicates that the shapes of migration age profiles are much more stable than their levels.[7] Data such as those in Figure 1 therefore present the researcher with an opportunity. The discrepancies between lifetime fractions and accumulated current rates contain information about changes in rates over cohorts' lifetimes.[8] Furthermore, although many other factors also influence lifetime fractions, it is probable that the gaps in Figure 1 are due primarily to changes in the overall level of migration over time – i.e. to period effects. Estimates of these period effects would be interesting measures of historical migration trends. In the next section a model and estimation procedure for this purpose is outlined.

Model

In order to focus on period effects, assume that migration hazards consist of a varying period-specific component and a fixed age-specific component. Write this as:[9]

$$m_x(r,s,p) = \gamma_{rs}(p)\, m_x(r,s), \quad r \neq s. \tag{6}$$

Suppose also that the period-specific component γ follows a time trend with a simple parametric form:

$$\gamma_{rs}(p) = \exp\,[a_{rs}(p') + b_{rs}(p')^2], \quad r \neq s \tag{7}$$

where

$$p' = \left[\frac{p-\tau}{n}\right] \tag{8}$$

is a simple transformation of periods into n-year units for computational convenience. This parameterization allows a substantial variety of trends, depending on the signs and magnitudes of the a and b terms.[10]

[6] Because return migration is a relatively rare event, the lifetime fraction curves in Figure 1 may be taken as good approximations to 'proportion ever migrated', which is in turn approximately equal to the simple sum of hazards up to a given age. Thus, changes in the level of migration rates (age structure constant) would appear as increases or decreases in the proportion migrated in the oldest age group (75–79); changes in the age structure of rates (level constant) would appear as twists in the curve, with the proportion migrated in the highest age group remaining constant.

[7] See A. Rogers *Migration, Urbanization and Spatial Population Dynamics* (Boulder, Colorado, 1984) on general regularities in age-specific profiles. Some limited evidence on the stability of profile shapes over time can be found in A. Otomo and T. Itoh, 'Migration of the elderly in Japan'. In A. Rogers and W. Serow (eds.), *Elderly Migration: An International Comparative Study* (Boulder, Colorado, 1988); D. Vergoosen and F. Willekens, 'Migration of the elderly in the Netherlands', *ibid.*; P. Korcelli and Alina Potrykowska, 'Migration of the elderly in Poland', *ibid.* for Japan, the Netherlands, and Poland, respectively.

[8] This is similar in spirit to Brass's P/F fertility method (W. Brass and A. J. Coale, 'Method of analysis and estimation'. In W. Brass *et al.* (eds.), *The Demography of Tropical Africa* (Princeton, 1968)), which is also based on the comparison of lifetime rates (parity) with accumulated current rates (fertility). Brass assumed that rates have remained constant, and that discrepancies are due to reporting and measurement errors. In contrast, I assume that both current and lifetime rates are correctly reported, and that discrepancies are due to changes in rates over time.

[9] As noted above, the $r = s$ hazards are defined implicitly.

[10] There are many reasonable parameterizations of γ. The basic functional form chosen here – an exponential function of a polynomial in p' – has several advantages: it always yields positive hazards, it always yields $\gamma(\tau) = 1$, and it allows major reversals in trends over time (one reversal in the case of two polynomial terms, two reversals in the case of three terms, etc.). For the country and period studied (Brazil, 1900–80), allowing a single reversal in the time trend seemed appropriate; hence the specific form in the above equation.

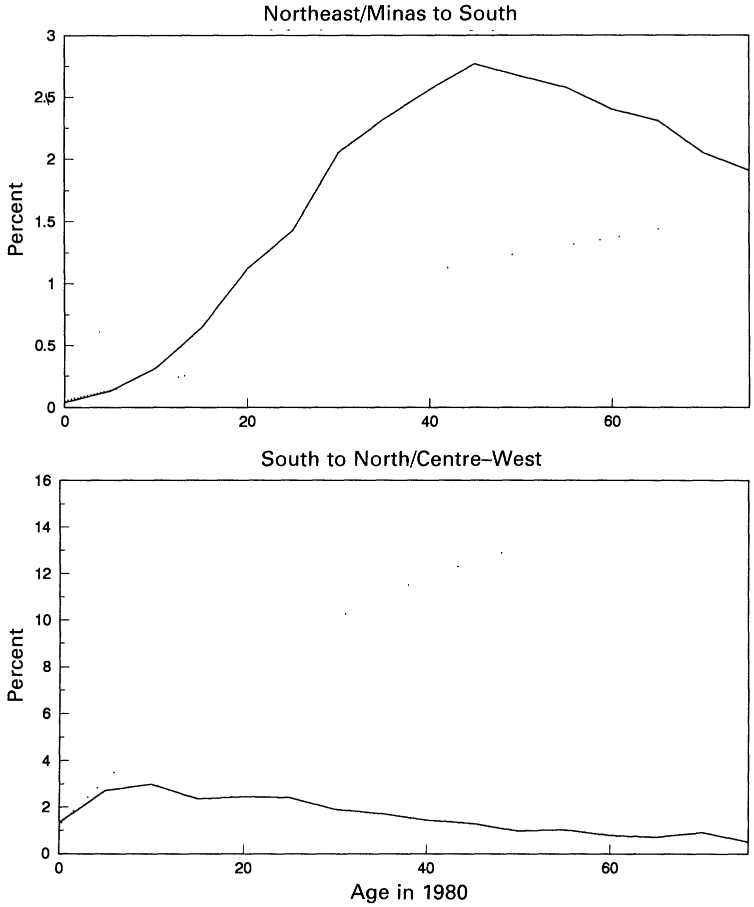

Figure 1. Lifetime migration fractions observed in 1980 compared with those expected if 1975–80 period migration rates had obtained over cohorts' lifetimes. ——, observed fractions; ····, expected fractions at 1975–80 rates.

Using this parametric framework, it now becomes possible to set up a regression model to estimate historical migration trends. For every set $(\mathbf{a},\mathbf{b})$ of $2R(R-1)$ parameters

$$(\mathbf{a},\mathbf{b}) = \{a_{rs}, b_{rs};\ r = 1 \ldots R,\ s = 1 \ldots R,\quad r \neq s\} \tag{9}$$

transition probabilities $\hat{\mathbf{P}}_x(p)$ may be estimated from Equations (7), (6), and (1), and estimated lifetime fractions at the census date $\hat{\mathbf{F}}_x$ calculated from Equation (3):

$$\hat{\mathbf{F}}_x = \hat{\mathbf{P}}_x(\tau)\,\hat{\mathbf{P}}_{x-n}(\tau-n)\ldots\hat{\mathbf{P}}_0(\tau-x). \tag{10}$$

Each $(\mathbf{a}, \mathbf{b})$ thus implies a set of CR^2 estimated lifetime fractions – one for each (cohort, origin, destination) combination. Given several very benign assumptions about hazards and sampling errors, minimizing the sum of squared differences between these estimated lifetime fractions and those observed in the survey will yield statistically consistent estimators for $(\mathbf{a},\mathbf{b})$.[11]

[11] This is a non-linear least-squares problem, which requires estimation of $(\mathbf{a},\mathbf{b})$ by some iterative method. For the calculations in this paper I used Gauss–Newton iterative regression. See G. R. Judge, R. Carter-Hill, W. E. Griffith, H. Lutkepohl and Tsoung Chao-Lee, *Introduction to the Theory and Practice of Econometrics* (2nd ed., New York, 1988), pp. 489–511 for details.

Organizing the data during the estimation process is fairly simple. When calculating estimated **F**'s, the matrix framework above, with C distinct $R \times R$ matrices is used. For calculation of estimation errors and parameter updates, however, all the **F** and estimated **F** data must be re-arranged into $(CR^2 \times 1)$ vectors.

IV. APPLICATION TO BRAZILIAN DATA

Data and definitions

For this study I divided Brazil into four large regions: (1) the North and Centre–West (abbreviated NC throughout), (2) the Northeast, Minas Gerais, and Espírito Santo (NM), (3) the states of Rio de Janeiro, São Paulo (RS), and (4) the South (SO).[12] Data from the three per cent public use sample of the 1980 Brazilian Demographic Census, combined with a special tape from the Census Bureau which contains information from the 25 per cent national sample on all recent migrants, were used to construct 4×4 matrices $\mathbf{M}_x$ and $\mathbf{F}_x$ for each of 16 closed five-year birth cohorts, ranging from 0–4 to 75–79 years of age at the census date.[13] These matrices appear in tabular form in Tables A 1 and A 2 in the Appendix.

Parameter estimates

With 4 regions and 16 cohorts, the model outlined above has of 256 observations $[\mathbf{F}_0 \ldots \mathbf{F}_{75}]$, 256 'independent variables' $[\mathbf{M}_0 \ldots \mathbf{M}_{75}]$, and 24 parameters $(\mathbf{a}, \mathbf{b})$.[14] Least-squares parameter estimates appear in Appendix (Table A 3).

Based on asymptotic distributions, 16 of the 24 individual parameter estimates are significantly different from zero at the five per cent level, and for ten of the 12 interregional flows the hypothesis of constant rates over the period 1900–80 – i.e. that a and b are both zero – can be rejected at the one per cent level.[15] The most notable exception to the general pattern of significance is for the North/Centre–West to South flow. In this case the model fails because nearly 100 per cent of 'historical' migration in this direction seems to have taken place in the period immediately preceding the census (1975–80), a pattern which can be reproduced by an extremely wide range of (a_{rs}, b_{rs}) pairs.

Estimated lifetime fractions

The period-effects model fits the observed lifetime fractions well for all twelve interregional flows.[16] Simple trends in migration levels can account for almost all of the discrepancies between observed lifetime fractions and accumulated rates for 1975–80.

[12] For the Brazilian data used throughout this paper τ = Sept. 1980, $n = 5$, and there are 16 closed cohorts: $c = 0$ (0–4 years old at time τ) to $c = 15$ (75–79 at τ). Regions are defined as the following aggregations of States and territories.

North/Centre–West: (Rondônia, Acre, Amazonas, Roraima, Pará, Amapá, Mato Grosso, Mato Grosso do Sul, Goiás, Distrito Federal).

Northeast/Minas Gerais: (Maranhão, Piauí, Ceará, Rio Grande do Norte, Paraíba, Pernambuco, Alagoas, Fernando de Noronha, Sergipe, Bahia, Minas Gerais, Espírito Santo).

Rio/São Paulo: (Rio de Janeiro, São Paulo).

South: (Paraná, Santa Caterina, Rio Grande do Sul).

[13] Brazilian migration data contain information on place of last residence and duration of residence, rather than on place of residence at a fixed point in the past. Construction of the **M** matrices in this case requires a special procedure. Details are available from the author.

[14] Although estimation of such a large non-linear model appears formidable, it is well within the capacity of a standard (*c.* 1989) personal computer.

[15] Under the null-hypothesis of constant rates for each flow, $H_0: (a_{rs} = b_{rs} = 0)$, the values in the last column of Table A 3 are drawn from a distribution which is asymptotically $F(2, 232)$, and the critical values are 3.04 and 4.70, respectively, at the five per cent and one per cent levels.

[16] There are in fact 16 flows, since the fitting procedure also takes into account estimated fractions of non-movers (i.e. diagonal elements of **F**).

Figures 2 and 3 demonstrate this graphically: for each origin-destination pair, the solid line represents observed lifetime fractions, the dotted line represents expected lifetime fractions under the constant-rate hypothesis $(\mathbf{a}, \mathbf{b}) = 0$, and the dashed line represents fitted fractions from the period-effects model. The close fits obtained with the period-effects model suggest that the simplifying assumptions necessary for estimation are not excessively restrictive.

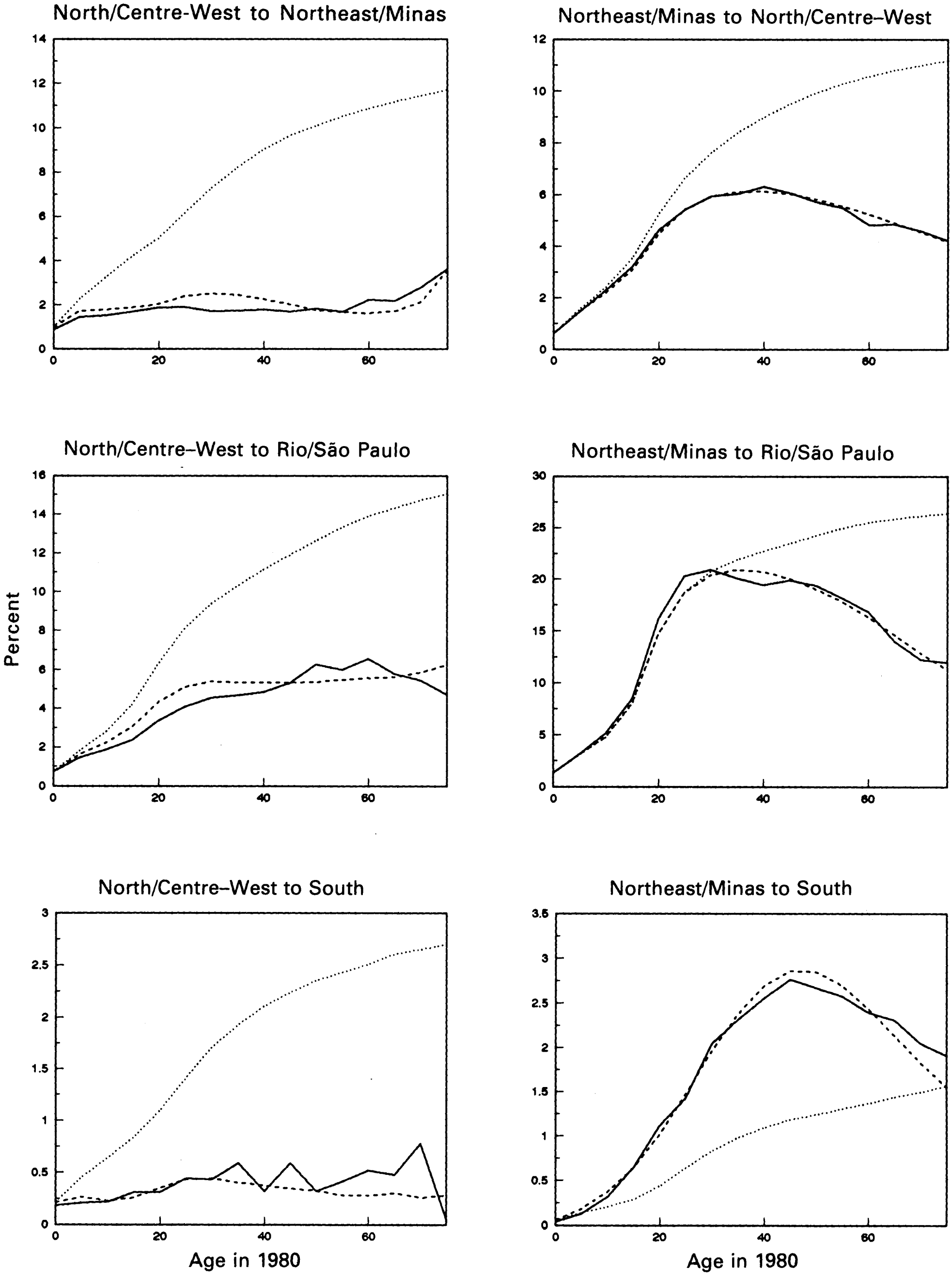

Figure 2. Lifetime migration fractions for various origin destination pairs. ——, Observed; ·····, expected at constant 1975–80 rates; -----, fitted fractions from period-effects model.

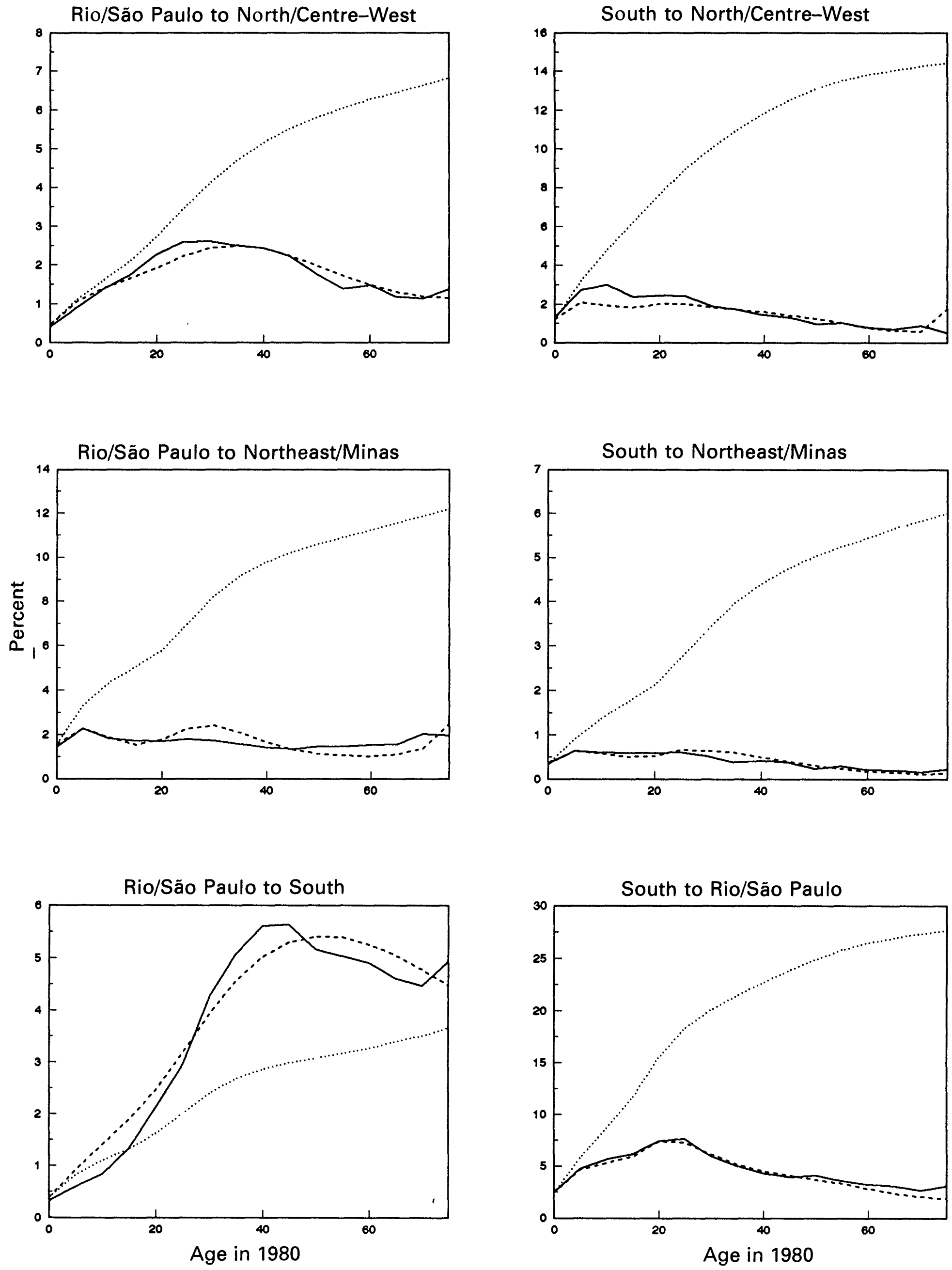

Figure 3. Lifetime migration fractions for various origin destination pairs. ——, Observed; ·····, expected at constant 1975–80 rates; -----, fitted fractions from period-effects model.

Estimated period effects

Figure 4 presents the estimated period effects in graphical form. Each of the four graphs corresponds to a different destination region, with trends representing estimated levels – relative to 1975–80 – of emigration rates from the other three regions to that destination.

Because trends for all 12 flows are normalized to unity for the period 1975–80, they allow comparison between periods, but not between origin-destination pairs. Nonetheless, several of the major findings from previous studies of Brazilian internal migration[17] are immediately apparent. For example, there is a general increase in interregional mobility over the course of the century, a rapid increase of migration to the South during the 1940s and 1950s followed by an equally rapid decrease in the 1960s, and an increasingly important role of emigration from the South, beginning in the 1960s.

Estimated gross flows 1900–80

Even readers familiar with Brazil may have difficulty in analysing the estimated period effects in Figure 4. One typically has some knowledge or intuition about the volume of historical migratory flows, rather than about relative migration rates over time. It would be much more informative to investigate the implications of the estimated period effects for the volume of migration flows between the various pairs of origins and destinations in each period. This requires the use of supplemental data, since the single-census method allows inferences about historical rates but provides no information about the sizes of the past populations at risk of migration.

Keeping with the spirit of the single-census method, one would like to minimize the use of data from outside sources in the estimation of historical flows. To this end, I suggest that it is possible to arrive at reasonable approximations to historical flows under two rather naive demographic assumptions: (1) the age structure of each region has remained approximately constant over the entire period under study; and (2) regional populations have grown at constant geometric rates between censuses. Given these assumptions, the matrix of total gross flows in the period $(t-n, t)$ can be approximated by[18]

$$\widehat{\mathbf{FLOW}}(t) = \sum_x \hat{\mathbf{P}}_x(t)\,\mathrm{DIAG}\{[\hat{\mathbf{P}}_x(t)]^{-1}\mathbf{N}(x,\tau)\,\hat{\mathbf{K}}(t)\} \tag{11}$$

where $\hat{\mathbf{P}}_x(t)$ is a transition matrix estimated from the period-effects model, $\mathbf{N}(x,\tau)$ is a diagonal matrix with its rth diagonal element equal to the fraction of region r's population in age group x at the time of the survey, $\hat{\mathbf{K}}(t)$ is the vector of total regional populations at time t (estimated by interpolation from known regional population totals in supplemental sources using the assumption of geometric growth), and DIAG is an operator which converts an $R \times 1$ vector $\mathbf{v}$ into an $R \times R$ diagonal matrix with $\mathbf{v}$ on the main diagonal.

Under this scheme the only data required from outside sources are total regional populations at several points in time. In the case of Brazil such estimates are available from the censuses of 1900, 1920, 1940, 1950, 1960, and 1970. These data appear in Table A 4 in the Appendix, together with the 1980 regional age distributions.

Despite its simplistic nature, analysis based on Equation (11) leads to estimates of historical migration flows which are qualitatively consistent with previous studies. In

[17] Merrick and Graham, and Graham and Hollanda have covered the period 1900–70. Carvalho has analysed interregional migration between 1940 and 1970. See T. W. Merrick and D. H. Graham, *Population and Economic Development in Brazil: 1800 to the Present* (Baltimore, 1979); D. H. Graham and S. Buarque de Hollanda Filho, *Migrações Internas no Brasil: 1872–1970* (São Paulo, 1984); J. Alberto Magna de Carvalho, 'Analysis of Regional Trends in Fertility, Mortality and Migration in Brazil, 1940–70', (unpublished Ph.D. dissertation, University of London, 1973).

[18] One can better understand this equation by considering the meaning of its various parts: $\mathbf{N}(x,\tau)\,\mathbf{K}(t)$ is simply the vector of regional populations in age group x at time t; premultiplying by $\mathbf{P}_x(t)^{-1}$ projects this population of survivors to t back to time $t-n$; converting the resulting vector into a diagonal matrix makes it possible to disaggregate flows by origin when projecting the population forward again by premultiplication by $\mathbf{P}_x(t)$.

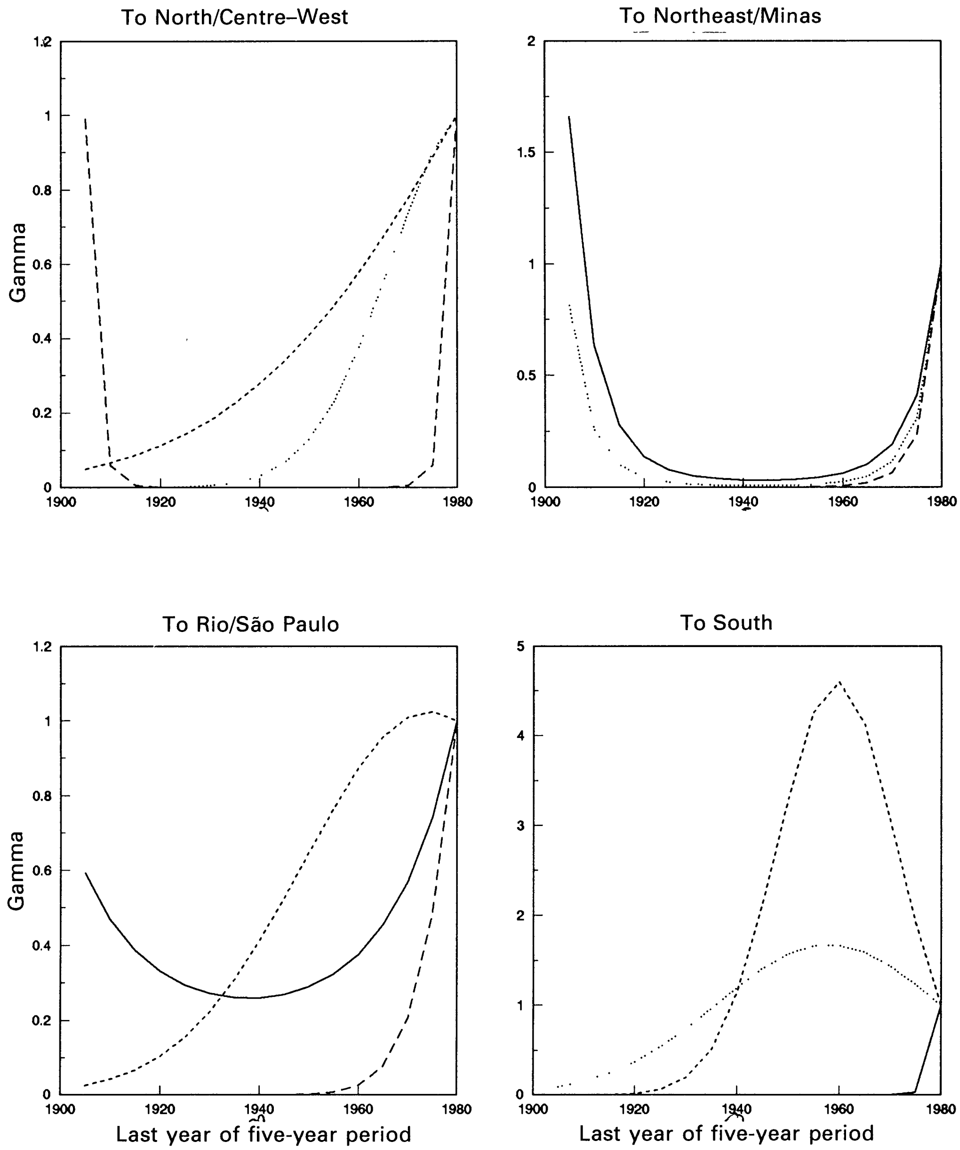

Figure 4. Estimated period effects for gross immigration to each region. ——, From North/Centre-West; -----, from Northeast/Minas; ·····, from Rio/São Paulo; —, from South.

Figures 5–7 the estimates obtained from Equation (11) are plotted in various forms. It is important to re-emphasize that the historical estimates in these graphs are derived almost exclusively from analysis of information in the Census of 1980.

In Figure 5 each of the four graphs corresponds to a destination, and the lines represent estimated gross inflows (by five-year period) from the other regions. In Figure 6 the resulting estimated total net inflows for each region in each period are plotted. In Figure 7 estimated national rates of interregional migration (gross and net, calculated as the total number of moves in the period[19] per person alive at the end of the period) are shown.

The addition of regional population totals to estimated period effects allows more

[19] Summing only over regions with positive net inflows when calculating total net moves.

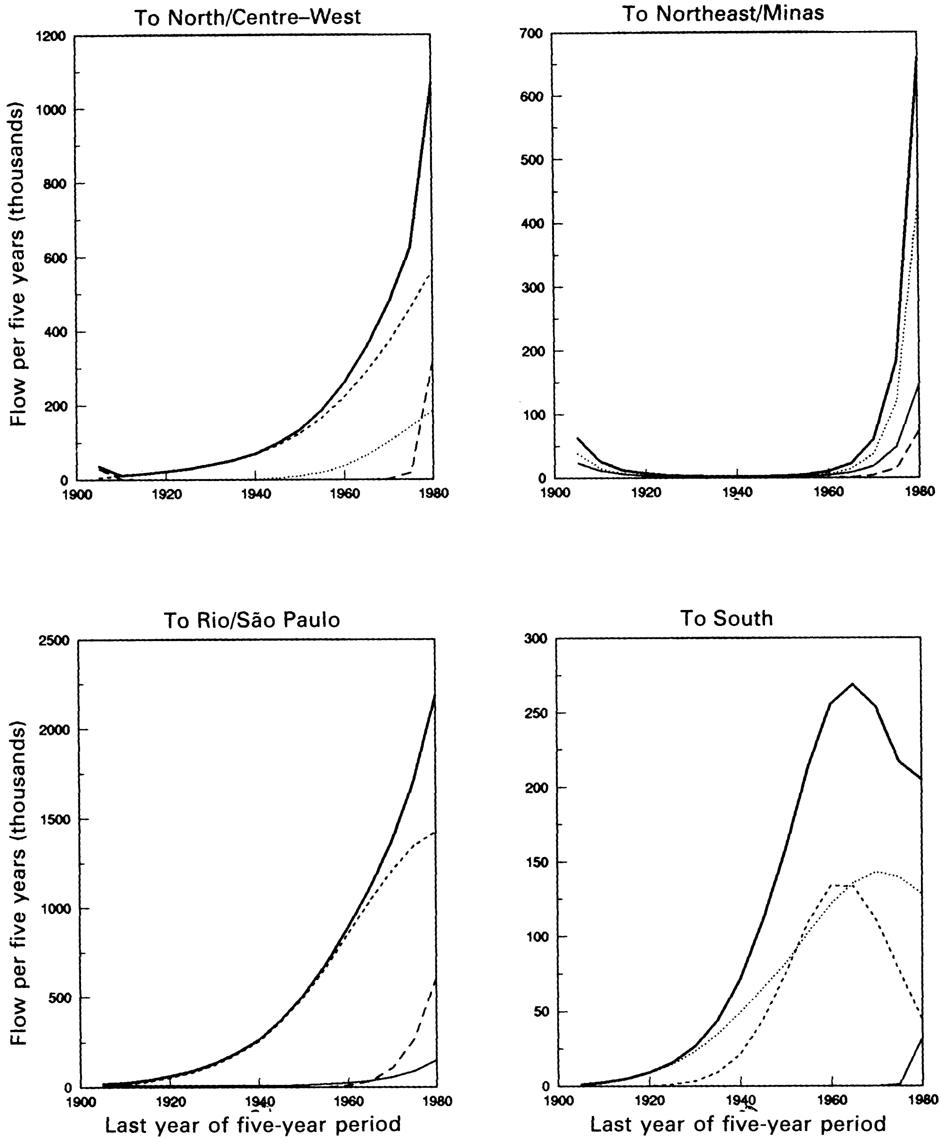

Figure 5. Estimated volume of gross immigration to each region. ——, From North/Centre–West; -----, from Northeast/Minas; ·····, from Rio/São Paulo; ––, from South; ——, total gross immigration.

detailed interpretation than was possible from Figure 4. For example, the data in Figures 5–7 indicate that the total volume of interregional migration in Brazil was very low until the 1920s and 1930s, despite estimates that several rates (SO to NC, NC to NM, RS to NM, and NC to RS) were relatively high early in the century. The finding of low levels of migration before the period 1920–40 is consistent with estimates from other sources.[20]

The estimates in Figure 5 also indicate that from 1920 to 1960 virtually all interregional migration in Brazil consisted of migration from the Northeast/Minas region, and that during this period the great bulk of those who left Northeast/Minas went to Rio/São Paulo. This is also consistent with previously published data.

[20] See Merrick and Graham, *op. cit.* in fn. 17, p. 124.

The peak in immigration to the South during the 1960s appears again, although in slightly different form, since the very sharp peak in relative NM to SO rates is damped, because these rates were never as high as those from RS to SO.

The fact that the period-effects model seems to 'discover' the same general patterns in historical flows found by standard methods is encouraging. In addition, the ability of the single-census method to estimate gross as well as net historical flows, combined with its ability to disaggregate these flows by origin, yields several new qualitative insights about Brazilian migration trends.

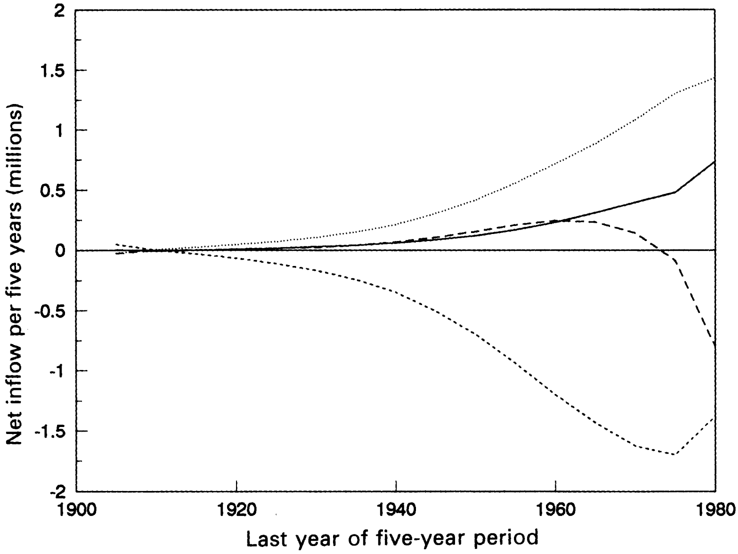

Figure 6. Estimated net immigration to each region. ——, To North/Centre–West; -----, to Northeast/Minas; ·····, to Rio/São Paulo; –-, to South.

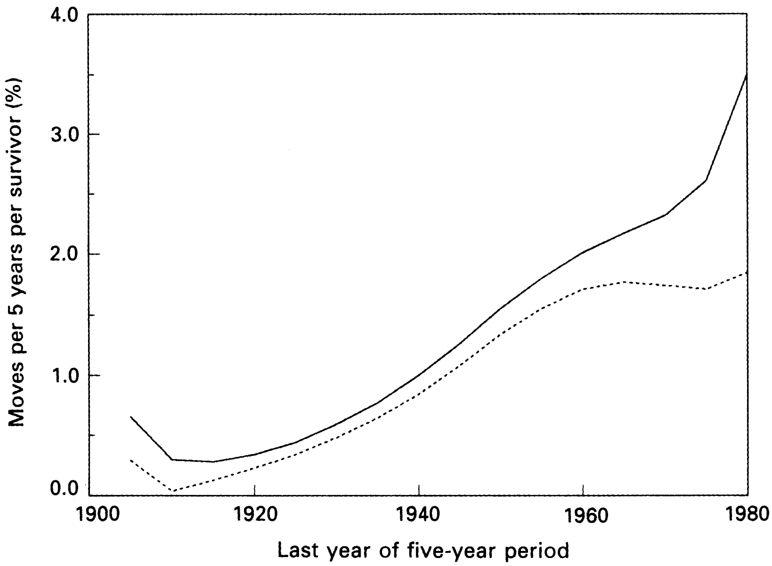

Figure 7. Estimated net and gross interregional migration for various five-year periods, measured as a percentage of national population at end of period. See text for definitions. ——, Gross interregional migration; -----, net interregional migration.

For example, it is possible to see several patterns which might be imperceptible in a standard analysis of net flows by region/rest of Brazil: (1) immigration into the North/Centre–West has occurred in three overlapping waves, first from the Northeast/Minas starting in the 1920s, second from Rio/São Paulo starting during the 1950s, and most recently a very large inflow from the South in the 1970s; (2) although net immigration to Northeast/Minas continues to be highly negative, gross inflows to the region increased significantly during the 1960s and especially the 1970s, enough to reverse the secular trend of increasingly large net outflows; (3) although net population transfer to Rio/São Paulo continued to increase, its composition changed during the 1960s and 1970s, with an increasing proportion of immigrants coming from outside Northeast/Minas, particularly from the South; (4) the sudden change from positive to negative net immigration experienced by the South in the 1970s was due almost exclusively to increased emigration from the region, rather than to decreased immigration, although immigration to the South from Northeast/Minas did decrease very significantly in percentage terms after a peak in the 1960s; and (5) previous findings of either decreasing[21] or decelerating[22] ratios of net migration to population for Brazil as a whole between 1950–60 and 1960–70 (also found here in Figure 7) should not be taken as evidence that interregional migration peaked during the 1950s; on the contrary, the model indicates that the number of interregional moves probably increased significantly from the 1950s to the 1960s, and that any slowdown in net transfers was due to increased migration in counterflows such as RS to NC, RS to NM, and SO to RS.

Estimated net intercensal flows: comparison with multiple-census methods

In the preceding section it was shown that the single-census method, combined with rudimentary information about past populations, can arrive at qualitative conclusions about historical migration which are very similar to those found by multiple-census methods, and that it can also yield important new information by disaggregating net into gross flows.

Before attempting any detailed interpretation or comparison of trends such as those in Figures 5–7, however, one would like to be confident that the estimated flows from Equation (11) are also quantitatively sound. In the case at hand (Brazil 1900–1980), data on net inflows to each region are available from two sources which used traditional multiple-census methods: Carvalho, and Graham and Hollanda.[23] Comparison confirms that application of the period-effects model to 1980 Brazilian data yields estimated net flows which are quantitatively similar to their estimates.

In Table A 5 the gross flows calculated from Equation (11) are aggregated into intercensal net flows for 1900–20, 1920–40, 1940–50, 1950–60, and 1960–70, and compared with Carvalho's and to Graham and Hollanda's numbers.[24] The same data are shown graphically in Figure 8. The three sets of estimates are clearly quite similar in absolute terms, although the reader should take considerable care in making comparisons, since vertical scales vary between graphs and since each of the three estimation methods covers slightly different populations.[25] Despite these difficulties, it is clear that combining estimated period effects with intelligent guesses about past regional

[21] Merrick and Graham, *op. cit.* in fn. 17, p. 124.
[22] Carvalho, *op. cit.* in fn. 17, p. 152.
[23] Cited in fn. 17.
[24] Where the two sets of multiple-census estimates overlap, Carvalho's are likely to be more reliable, since he accounted carefully for mortality differences by age and region. Graham and Hollanda's series, however, has the advantage of spanning the entire period of time covered by the single-census estimates. See Merrick and Graham for another comparison of the two multiple-census estimates, *op. cit.* in fn. 17, pp. 128–131.
[25] See the notes in Table A 5.

populations using Equation (11) can lead to net migration estimates which are quantitatively sound. This finding lends additional credibility to the interpretation of estimated period effects and trends in the previous section.

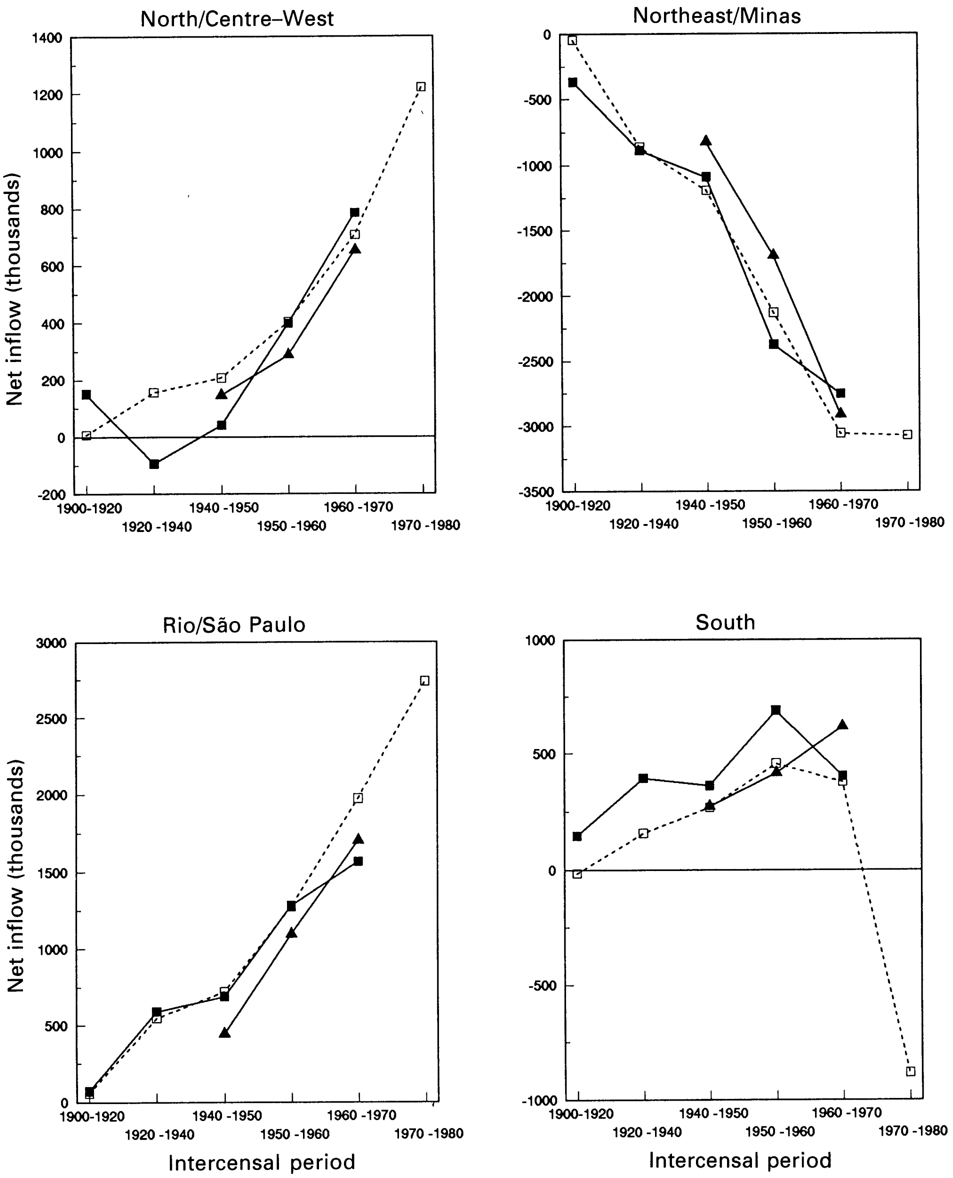

Figure 8. Estimated net intercensal migration to each region, comparing the period-effects model with existing indirect, multiple-census estimates. ■—■, Graham and Hollanda; ▲—▲, Carvalho; □---□, period-effects model.

V. CONCLUSION

The key to estimation of historical migration trends from data in a single census is the logical relationship between cohort and period 'migration tables' – i.e. between observed lifetime migration fractions and accumulated current rates. If there are no significant mortality differences between regions, then under a regime of unchanging migration rates these two measures will be nearly identical at every age.

Because rates are not in fact constant, different cohorts have experienced different rates during their lifetimes. The lifetime fractions of cohorts contain a considerable amount of information about these migration histories. Application to Brazilian data demonstrates that the researcher can recover a good deal of this information by comparing observed lifetime fractions to estimated lifetime fractions under different historical trends.

It is important to note that this argument applies not only to migration, but also to many other phenomena in both single-region and multiregional demography. The estimation method presented in this paper can, in principle, be applied to any situation in which the researcher can observe the proportions of each cohort in various states at three points in time. However, it exploits two features specific to migration, which may not obtain in other cases: the relative stability of the age structure of rates relative to the overall level of these rates, and the fact that realistic mortality differences between those in the various states represent a small influence on lifetime fractions relative to rate changes.

The single-census method is intended as a complement to traditional methods, rather than as a substitute. In cases where multiple surveys do not exist, of course, the single-census approach may be the only alternative. In general, however, one should probably consider multiple-census estimates as yielding more reliable numbers, particularly for small regions. The period-effects model then serves as a tool for disaggregating net into gross flows. As the above analysis has shown, this disaggregation can lead to interesting insights about the composition and timing of historical migration flows.

It is possible that applying the method to Brazilian data is an especially easy test, in the sense that the large changes and reversals in interregional migration patterns experienced in Brazil can be captured even by unrefined methods. Application to a less volatile case, such as the United States, would be an interesting second test. Experiments with different functional forms for period effects γ would also be useful, as would more careful accounting for interregional mortality differences. Another promising avenue of research would be to use model schedules for age-specific rates;[26] this would allow for time trends in parameters affecting the age structure of rates, as well as their overall level.

Despite the need for refinement, the single-census method performs very well in this first application, to interregional migration in twentieth-century Brazil: simple trends in period effects explain virtually 100 per cent of the discrepancies between lifetime fractions and accumulated current rates; the resulting estimates of historical trends and migration flows are both qualitatively and quantitatively consistent with previous studies; and the dissaggregation of estimates into gross rather than net flows leads to insights which are not possible with standard methods.

[26] See Rogers, *op. cit.* in fn. 7.

APPENDIX TABLES

Table A 1. *Annual hazard rates (per 1000) 1975–80*

Cohort (age in 1980)	From NC to			From NM to			From RS to			From SO to		
	NM	RS	SO	NC	RS	SO	NC	NM	SO	NC	NM	RS
0–4	4.03	2.99	0.87	2.49	5.42	0.21	1.86	6.49	1.58	5.07	1.30	10.06
5–9	2.74	2.24	0.53	2.00	3.82	0.16	1.29	3.71	0.98	4.25	1.06	7.49
10–14	2.23	2.05	0.44	1.88	3.52	0.12	1.02	2.40	0.70	3.79	0.82	6.50
15–19	2.36	2.78	0.50	2.63	7.34	0.16	1.01	1.97	0.62	3.61	0.69	7.39
20–24	2.74	4.06	0.68	4.24	16.56	0.27	1.27	2.81	0.85	4.05	0.75	10.08
25–29	3.27	3.77	0.81	3.85	11.52	0.31	1.53	3.72	1.12	4.03	0.97	8.45
30–34	3.11	3.10	0.78	3.00	6.93	0.26	1.48	3.62	1.11	3.64	0.88	6.25
35–39	2.74	2.34	0.64	2.38	4.38	0.20	1.28	2.75	0.90	3.34	0.84	5.22
40–44	2.36	2.18	0.56	2.05	3.35	0.17	1.12	2.00	0.69	3.09	0.67	4.76
45–49	1.95	1.97	0.47	1.77	2.99	0.16	0.88	1.54	0.51	2.71	0.54	4.39
50–54	1.53	1.83	0.42	1.56	2.88	0.09	0.69	1.34	0.43	2.42	0.39	3.91
55–59	1.48	1.78	0.30	1.37	2.75	0.10	0.60	1.27	0.43	1.95	0.30	3.59
60–64	1.24	1.56	0.28	1.16	2.32	0.08	0.56	1.16	0.37	1.44	0.21	2.75
65–69	1.06	1.20	0.29	1.00	1.68	0.08	0.48	1.18	0.47	1.15	0.21	2.37
70–74	0.88	1.24	0.16	0.82	1.41	0.06	0.52	1.03	0.42	0.96	0.15	2.11
75–79	0.87	1.08	0.13	0.73	1.46	0.05	0.51	1.13	0.50	1.02	0.10	1.85

Table A 2. *Lifetime migration fractions (per cent) 1980*

Cohort (age in 1980)	From NC to			From NM to			From RS to			From SO to		
	NM	RS	SO	NC	RS	SO	NC	NM	SO	NC	NM	RS
0–4	0.87	0.72	0.18	0.62	1.34	0.04	0.41	1.44	0.32	1.34	0.36	2.49
5–9	1.45	1.48	0.21	1.47	3.16	0.13	0.90	2.27	0.60	2.72	0.64	4.78
10–14	1.53	1.88	0.22	2.31	5.13	0.32	1.38	1.83	0.83	2.99	0.61	5.63
15–19	1.69	2.37	0.31	3.22	8.47	0.65	1.75	1.74	1.33	2.37	0.59	6.18
20–24	1.88	3.38	0.31	4.64	16.20	1.12	2.27	1.72	2.13	2.46	0.59	7.43
25–29	1.91	4.09	0.44	5.41	20.35	1.43	2.60	1.81	2.96	2.42	0.61	7.63
30–34	1.72	4.55	0.43	5.95	20.95	2.05	2.61	1.74	4.26	1.92	0.52	5.94
35–39	1.73	4.67	0.59	6.02	20.13	2.32	2.51	1.57	5.07	1.73	0.38	5.02
40–44	1.78	4.85	0.32	6.32	19.44	2.56	2.44	1.43	5.61	1.45	0.42	4.31
45–49	1.70	5.31	0.59	6.06	19.90	2.77	2.23	1.35	5.63	1.31	0.38	3.96
50–54	1.83	6.27	0.32	5.71	19.41	2.67	1.77	1.48	5.17	0.99	0.24	4.10
55–59	1.70	5.99	0.41	5.50	18.18	2.58	1.39	1.46	5.04	1.04	0.30	3.61
60–64	2.24	6.57	0.52	4.84	16.83	2.40	1.49	1.55	4.91	0.79	0.21	3.25
65–69	2.19	5.77	0.48	4.86	14.01	2.31	1.18	1.57	4.61	0.72	0.20	3.07
70–74	2.82	5.44	0.78	4.61	12.27	2.05	1.14	2.04	4.47	0.91	0.17	2.68
75–79	3.66	4.70	0.05	4.22	11.99	1.91	1.37	1.98	4.94	0.52	0.23	3.08

Table A 3. *Parameter estimates*

	a Estimate	a S.E.	a t	b Estimate	b S.E.	b t	F stat for H0: $(a, b) = (0, 0)$
From NC to							
NM	0.9567	0.0785	12.1891**	0.0660	0.0052	12.6981**	84.38**
RS	0.3204	0.0247	12.9976**	0.0190	0.0027	7.1541**	334.80**
SO	3.7756	625.6235	0.0060	0.1890	565.0559	0.0003	0.02
From NM to							
NC	0.1153	0.0322	3.5804**	−0.0056	0.0048	−1.1615	216.00**
RS	−0.0436	0.0092	−4.7639**	−0.0193	0.0014	−13.6197**	744.07**
SO	−0.7461	0.1484	−5.0265**	−0.0912	0.0286	−3.1933**	125.69**
From RS to							
NC	0.0587	0.2057	0.2852	−0.0471	0.0545	−0.8639	21.63**
NM	1.2474	0.1423	8.7660**	0.0822	0.0086	9.5517**	61.55**
SO	−0.2343	0.0358	−6.5524**	−0.0265	0.0051	−5.1660**	52.56**
From SO to							
NC	3.0266	1.2118	2.4977*	0.2017	0.0523	3.8579**	1849.83**
NM	1.5098	3.0501	0.4950	0.0685	1.4804	0.0463	1.12
RS	0.6613	0.2752	2.4034*	−0.0649	0.1334	−0.4864	203.43**
Number of observations				256			
Number of parameters				24			
Degrees of freedom				232			
Sum of squares [F-accum current rate]				30743.73			
Sum of squared prediction errors				72.91			
Pseudo-R^2				0.998			

Notes

(1) Dependent variable and errors measured in percentages.

(2) Standard errors in parentheses based on asymptotic (i.e. approximate) distributions.

(3) (*, **) = estimate significantly different from zero at five per cent and one per cent level, respectively, based on asymptotic tests.

(4) Pseudo-R^2 is a measure of model's success at explaining discrepancies between observed lifetime fractions F and lifetime fractions predicted under regime of unchanging rates. It is calculated as the difference between unity and the ratio of sum of squared prediction errors to sum of squared discrepancies.

Table A 4. *Regional populations and age distributions*

	NC	NM	RS	SO
	Total population (000s)			
Year				
1980	13,338	50,068	35,353	18,867
1970	8,682	41,235	26,770	16,518
1960	5,489	33,664	19,427	11,540
1950	3,582	26,713	13,809	7,841
1940	2,721	21,988	10,792	5,735
1920	2,122	17,427	6,167	3,282
1900	1,005	9,899	2,979	1,475
	1980 Age distributions			
Age				
0–4	0.1629	0.1528	0.1216	0.1255
5–9	0.1416	0.1350	0.1069	0.1179
10–14	0.1300	0.1312	0.1033	0.1213
15–19	0.1153	0.1170	0.1094	0.1198
20–24	0.0963	0.0892	0.1076	0.1003
25–29	0.0779	0.0698	0.0933	0.0825
30–34	0.0617	0.0577	0.0742	0.0675
35–39	0.0500	0.0485	0.0596	0.0551
40–44	0.0438	0.0447	0.0531	0.0493
45–49	0.0330	0.0351	0.0441	0.0410
50–54	0.0268	0.0317	0.0393	0.0352
55–59	0.0196	0.0249	0.0290	0.0269
60–64	0.0145	0.0198	0.0219	0.0212
65–69	0.0123	0.0178	0.0164	0.0160
70–74	0.0075	0.0120	0.0101	0.0100
75–79	0.0043	0.0076	0.0060	0.0060

Note

(1) 1900–20 populations from Graham and Hollanda *op. cit.* in fn. 17; 1940–70 populations from Carvalho *op. cit.* in fn. 17; 1980 populations from three per cent public use sample.

Table A 5. *Estimated intercensal net immigration (1000s)*

	1900 –1920	1920 –1940	1940 –1950	1950 –1960	1960 –1970	1970 –1980
To North/Centre-West						
Graham and Hollanda (1984)	151	−94	41	398	785	NA
Carvalho (1973)	NA	NA	146	287	653	NA
Model	7	156	207	402	708	1220
To Northeast/Minas						
Graham and Hollanda (1984)	−367	−888	−1090	−2376	−2753	NA
Carvalho (1973)	NA	NA	−823	−1695	−2914	NA
Model	−48	−861	−1194	−2132	−3058	−3075
To Rio/São Paulo						
Graham and Hollanda (1984)	70	590	689	1281	1568	NA
Carvalho (1973)	NA	NA	444	1097	1703	NA
Model	56	548	720	1274	1973	2739
To South						
Graham and Hollanda (1984)	146	392	360	687	401	NA
Carvalho (1973)	NA	NA	272	415	617	NA
Model	−15	157	267	456	377	−884

Notes

(1) Graham and Hollanda estimates are for net immigration among population 0+ years old at first census, and surviving to second census.

(2) Carvalho estimates are for net immigration among population 0–49 years old at first census, and surviving to second census.

(3) Model estimates are for net immigration among entire population alive at second census date.