Daniel COURGEAU,* Salut MUHIDIN,** Martin BELL***

Estimating Changes of Residence for Cross-National Comparison

The problems of temporal and spatial comparability of internal migration are linked to the ways it is measured and have been studied since the end of the nineteenth century (Ravenstein, 1885, 1889; Lee, 1966). They have been addressed through a number of models which attempt to link the count of migrants or migrations to the period of observation (Myers et al., 1967; Ginsberg, 1972), and to the physical distance covered (Zipf, 1946; Hägerstrand, 1957). Some authors have employed a more sociological approach to measuring distance, as in the intervening opportunities model proposed by Stouffer (1940).

In 1973 Courgeau proposed a range of general models and parameters designed to summarize the effect of differing observation periods on internal migration rates (1973a; English version, 1979). In the same year (1973b) he examined the effect of space on migration rates and proposed an index capturing these effects. In practice, these models are closely related and can be applied to address the following questions: "Is the same temporal model appropriate for every kind of division of the territory and, if so, how do its parameters change with different levels of spatial disaggregation? Conversely, is the same spatial model appropriate for migration measured over different intervals and, if so, how do its parameters change?". The two papers set forward some general answers to these questions.

The temporal model has been re-examined recently with more detailed and later data. Donzeau and Pan Ké Shon (2009), for example, showed that only by adopting the question "where were you living a year ago" at the census, is it possible to conduct robust investigation of sub-populations.(1) In addition, Royer (2009) concludes by saying:

(1) Surprisingly, the one year question was asked in just 31 countries in the 2000 round of national censuses (Bell et al., 2011). In 2011 it was also the question adopted for the census in France.

* Institut national d'études démographiques, Paris.
** Faculty of Business and Economics, Macquarie University, Australia.
*** School of Geography, Planning and Environmental Management, University of Queensland, Australia.

Correspondence: Daniel Courgeau, Les Coccinelles, 114-2 chemin du Défends, 06250 Mougins, e-mail: daniel.courgeau@wanadoo.fr

> His [Courgeau, 1973a] results stand up remarkably well and his conclusions are largely confirmed, both for the shape of the migrant-migration relationship over time within the study period and for the order of magnitude of the main coefficients in this relationship, except for return migration.

The spatial model has been re-examined in the context of developing a suite of comparative measures of internal migration (Bell et al., 2002) and extended to a larger selection of countries in Europe (Rees and Kupiszewski, 1999) and worldwide (Bell and Muhidin, 2009, and 2011). These analyses showed that it was possible to summarize the migration intensity of a large range of countries with the single index *k* and to compare their levels of mobility. Bell and Muhidin (2011) conclude:

> It [the index *k*] enables contrasts to be drawn between countries with quite different zonal geographies. When used for trend analysis it automatically corrects for some aspects of boundary changes. It is simple to compute and can be implemented with only a small number of observations: indeed, it even provides a graphical space in which to situate countries for which only a single observation is available.

However, they also pointed to several limitations:

> Most critical, perhaps, is that the index value *k* has no intrinsic or plain language meaning. Thus, *k* is not directly interpretable as a demographic indicator in the same form as the total fertility rate, life expectancy or migration intensity.

Courgeau (1973b, 1982) did not attempt to provide a precise definition of *k* or assess its international comparability, since the main focus was to examine the effects of population density and the geographic area on migration intensity. However, we will show here that the above criticism may be removed by coupling the index with a count of households and transforming the way we express the partitioning of space, to facilitate retrieval of the overall rate of residential mobility or change, which is the only internationally comparable index with a clear plain language meaning as advocated by Bell and Muhidin (2011).

We first summarize the hypotheses and main theoretical findings of the methods derived from Courgeau (1973b). We then examine their validation by empirical measurements, and show how *k* can be adapted to generate this more satisfying "plain language" index, before concluding.

I. Main hypotheses and theoretical findings of the method

These findings are the results first, of theoretical reasoning, and second, of the practical application of these findings to a number of countries.

The theoretical basis of this method is as follows: if, as we know to be true, there is a relationship between the propensity to move and distance, there must also be a relationship between the level of mobility and the number of zones dividing the territory. The problem is to determine this relationship.

A great number of migration models have been used to summarize the relationship between migration and distance (Courgeau, 1970). However, the model most often verified is a Pareto model defined as

$$m(r) = \frac{K}{r^{\alpha}} \qquad (1)$$

where $m(r)$ is the probability of migration between two areas[(2)] at a physical distance r, with K and α being the parameters to be estimated from the observed flows. Hägerstrand (1957) found values of α varying from 0.4 to 3.3, but the majority of values are close to 2.0. This model has been verified for a very large range of distances, extending from a hundred meters to thousands of kilometres (Jakobson, 1969). For these reasons we can initially set $\alpha = 2$, and modify the results with different values of α if required.

Like all geographical studies, major challenges for migration analysis arise from the great variety of shapes and sizes of the zones into which territories are divided. These issues are commonly recognized as the Modifiable Areal Unit Problem (MAUP) and hinge on the fact that the number of migrants or migrations recorded in any spatial system, and hence the apparent migration intensity, is fundamentally dependent on the number of zones in a country and the shape of those spatial units. These are identified as the scale and zonation dimensions of the MAUP (see Openshaw, 1984, Holt et al., 1996).

Here, we focus on the issue of spatial scale, and first examine what occurs when the size and dimensions of each zone are uniform and the territory of simple shape (square or equilateral triangle). In this instance, if one side of the territory is divided into n parts, the territory as a whole will be divided into n^2 zones. Under these theoretical conditions, it can be shown, (Courgeau, 1973b) that the crude migration intensity (CMI), measured by the number of migrants or migrations which cross at least one boundary between n^2 zones, divided by the total population at risk, is equal to

$$CMI = k \ln(n^2) = 2k \ln(n) \qquad (2)$$

where the parameter k is proportional to the density δ and the previous parameter K. This result holds for both square and triangular zones and appears to be independent of zonal shape.

The derivation of equation (2), originally presented in Courgeau (1973b, French version), is elaborated in Appendix 1. There, we also demonstrate refinements to equation (2) as two key assumptions are relaxed. The first concerns the considerable variation of population densities in a country. When we introduce a variation in density (δ) the equation becomes

(2) The areas in this instance are intended to be very small, so that their populations will not have to be taken into account in the model (each area containing only one individual).

$$CMI = k' \ln(n^2) + b \tag{3}$$

introducing a non-zero intercept, b, on the intensity axis, as n cannot be equal to 1. As discussed in the next section, this equation was estimated for a number of countries in Bell and Muhidin (2011).

The second refinement involves modifying the distance-decay parameter, α, in equation (1), originally set at 2.0, such that

$$m(r) = \frac{K}{r^{\alpha}}$$

Equation (2) then becomes

$$CMI = k''\left(1 - \frac{1}{n^{2-\alpha}}\right) f(\alpha) \tag{4}$$

where

$$k'' = k\frac{S}{a^{\alpha}} \tag{5}$$

These theoretical results show that the relationship between the level of mobility and the number of zones into which the space is divided seems to be quite simply related to the zoning of the territory. In the following section, we examine if these results are verified by empirical measurement of migrations in countries with different zonal systems. Prior to that analysis, however, it is also important to consider the temporal dimension of migration.

Migration can be measured in different ways, but the two most common approaches measure changes of residence either as transitions or as events. Migration events are typical of population registers which record all the internal migrations made by the observed population. Transitions, on the other hand, are commonly associated with censuses, which simply compare the places of residence of individuals at the beginning and end of a specified time interval. This is commonly one or five years, but some countries use other intervals (Bell et al., 2002). Transition data indicate the number of migrants, rather than the number of migrations.

Data on migrations present no difficulties for comparison, provided they are measured during the same general period. In contrast, data measuring the number of migrants need to be based on similar observation intervals, for example one year or five years,(3) to ensure comparability (Bell et al., 2002). As the observation interval lengthens, the data are increasingly affected by return and repeat migration, the incidence of which may vary between countries (Long and Bortlein, 1990; Rogerson, 1990). Is it possible to eliminate this effect and to determine a comparable instantaneous or annual(4) migration rate?

(3) For lifetime migrants no clear comparison is possible.

(4) If the best measure is an instantaneous migration rate, the use of an annual migration rate will be very similar. The numbers of migrants and migrations will be quite similar over a one-year period.

The *migrant-migration* model proposed by Courgeau (1973a), derived from the *mover-stayer* model (Blumen et al., 1955), permits this elaboration, based on fixed probabilities of repeat and return migration. Originally presented in French (Courgeau 1973a), the model is set out in detail in Appendix 2. The one-year/five-year problem has also attracted attention from other analysts (Kitsul and Philipov 1981, Rogers et al. 2003). Most recently, Nowok (2010) and Nowok and Willekens (2011) used probabilistic methods to chart the ratio of transitions measured over different intervals at various levels of migration intensity, in the context of the broader issue of harmonizing migration data on various temporal dimensions for European Union member states. While these contributions reveal the theoretical links and empirical parameters surrounding temporal measures of migration, there is no straightforward means to harmonize migration data measured over different intervals. In what follows, we therefore focus on data measured over similar intervals.

II. Validation of the spatial model by empirical measurement

The first test of the spatial model was made by Courgeau (1973b) for 11 countries around the world. The numbers of zones for which data were available ranged widely, from 37,962 to 22 zones in France for example, and from just 32 to 8 zones in Mexico. No correction was made for differences in the observation intervals for these measurements, which extended from one year (Japan, West Germany), to five years (Great Britain, United States) and to lifetime migration in India and Tunisia.

Under these conditions the test was not made to compare migration between countries, but only to verify the validity of the model under differences in zoning of the territories. The model was generally well verified. The countries giving the worst fit were India and Tunisia, which used lifetime migration.

In the same paper the author tried to eliminate differences on the temporal scale for three countries (France, Great Britain and the United States) in order to generate more comparable results. He used the migrant-migration model (Courgeau, 1973a) to harmonize the observation interval for the three countries to five years. The three countries ranked from low to high mobility in the same order, the index for the United States being 38% higher than that for France.

A more rigorous test was made by Courgeau (1982) comparing France (1968-1975) and the United States (1970-1980), in which instantaneous probabilities of migration were estimated, again using the migrant-migration model, with the same age structure for the two countries.(5) With those corrections, the mobility index was 64% higher in the United States than in France. Comparison with the previous result, even though the observation period was

(5) The effect of the correction for age structure was negligible.

not exactly the same (the previous data were from the 1960s), shows the importance of using instantaneous probabilities rather than five-year rates. [6]

Rees and Kupiszewski (1999) extended the application of equation (3) to compare migration intensities across ten European countries, using a mix of census and register data adjusted to single year intervals. In many cases data were available for only two spatial divisions. Nevertheless, they were able to identify distinctive differences between countries, with Norway, the Netherlands and Great Britain returning the highest values ($k' > 0.5$), followed by Portugal ($k' = 0.4$), then Estonia, the Czech Republic, Romania and Italy ($k' < 0.33$). Of the ten, only the Netherlands, Romania and Germany showed an increase in mobility over the decade 1984-1994, the latter consequent on unification in 1990 (Rees and Kupiszewski, 1999).

More recently, Bell and Muhidin (2011) undertook a comparison for 27 countries by using census data drawn mainly from the Integrated Public Use Microdata Series (IPUMS). They focused primarily on five-year migration intensities, since this is the fixed interval for which data are most commonly collected in censuses around the world, and compared the regression lines given by equation (3), in this case reporting on the *x*-axis the logarithm of the square of the number of regions. They estimated the parameter k computed when constraining the regression analysis to pass through the origin, set out the coefficient of determination R^2 as a measure of goodness of fit, and also reported the floating intercept calculated when the regression was unconstrained, for which they gave a possible explanation to which we return below. The results presented in Table 1 for the 17 countries with five-year transition data indicate a close linear relationship with most R^2 values varying between 0.89 and 0.99, though these were based on a small number of observations, corresponding to migrations between standard levels of statistical geography in each country (e.g. districts, counties, communes). Three countries – the Philippines, South Africa and the United States – gave a poorer fit, though, as will be shown below for the United States, a much better result is obtained using the 2000-2005 Current Population Surveys (CPS) data (R^2=0.9992), with a larger number of zonings.[7] The values of k in Table 1 are not directly comparable with those reported in the studies by Courgeau (1973b, 1982) and by Rees and Kupiszewski (1999) because they are based on data measured over different intervals. Nevertheless, Bell and Muhidin (2011) were able to clearly differentiate mobility levels among the mass of countries using the k' values.

(6) This is particularly significant because inter-communal annual mobility in France (Baccaïni et al., 1993) increased from 3.89% to 4.11%, during the 1962-1982 period, and residential mobility in the United States decreased slightly from 19.6% to 17.2%, during the 20 years 1961-1980 (Current Population Survey). The measurement during a longer period of five years reduces the differences observed for shorter periods, due to different parameters in the migrant-migration model.

(7) Migration data used in Bell and Muhidin (2011) were derived from censuses, including the 2000 census for USA. The zoning was restricted to 3 zonal systems: Region, Division and State, whereas the CPS data report five levels of spatial disaggregation.

The countries where the population was found to be most mobile were South Africa and Chile, followed closely by Costa Rica, the United States and Australia. Vietnam, the Philippines, Indonesia and China clearly emerged as countries with lower mobility.

Table 1. Courgeau's k and R^2 for selected countries, census five-year transitions

Country	Five years ended	k	R^2	Floating intercept
Africa				
Ghana	2000	1.320	0.903	1.176
South Africa	2002	3.006	0.819	–6.918
Asia				
China	2000	0.864	0.994	–2.100
Indonesia	2000	0.676	0.963	–0.621
Malaysia	2000	1.656	0.988	0.689
Philippines	2000	0.674	0.674	1.310
Vietnam	1999	0.806	0.890	–0.081
Latin America				
Argentina	2001	1.148	0.998	0.242
Brazil	2000	1.308	0.972	–0.556
Chile	2002	2.910	0.895	–4.456
Colombia	2005	1.060	0.942	1.107
Costa Rica	2000	2.536	0.964	1.303
Ecuador	2001	1.728	0.984	0.799
Developed countries				
Australia	2006	2.478	0.988	–1.004
Canada	2001	1.974	0.959	–1.042
Portugal	2001	1.514	0.904	–2.593
USA	2000	2.534	0.570	2.800

Note: Bell and Muhidin (2011) reported computed values of k based on the square of the number of zones. The values of k reported in this table have been recalculated based on the number of zones in place of the number of zones squared, to facilitate comparison with earlier work cited in the text.
Source: Modified after Bell and Muhidin (2011), and Current Population Survey (2005).

Bell and Muhidin (2011) also applied the model to lifetime migration and showed that the broad sequencing of geographic regions was maintained and the R^2 values remained strong, counter to the previous results obtained by Courgeau (1973b) for India and Tunisia. Applying the model to both five-year and lifetime migration over the previous 40 years provided a window on temporal trends, but showed some surprising results: a decline for 5-year migration intensities and an increase for lifetime migration for a majority of countries. They explain these differences as follows:

> Our interpretation of these disparate trajectories in the two measures is that contemporary trends in inter-regional migration are continuing to generate displacements in the pattern of human settlement throughout the world, as captured in the lifetime measure, but that these increases are occurring at a decreasing rate.

They concluded that a more realistic picture of contemporary trends is provided by examining intensities measured over a succession of fixed intervals, and ideally such comparisons should be based on data for a single year interval to eliminate the effect of unknown time parameters, which are excluded when using an instantaneous rate. Unfortunately the absence of the necessary data to apply a migrant-migration model did not allow them to eliminate this effect in the results presented here.

III. Linking the parameter *k* to the probability of a residential move

We have established from the theoretical model that the parameter k is linked to different characteristics of migration: to the parameter K of the Pareto model, to the exponent α of the same model, to the mean density of population of the country, which is more specifically proportional to the population of the country and conversely proportional to its area (Courgeau, 1973). This combination of parameters makes it difficult to provide a simpler exposition. Bell and Muhidin (2011) also concluded that the parameter k itself has no intrinsic plain language meaning. However, we can arrive at a practical solution by approaching the use of k in a different way.

We start with the observation that the probability of a residential move in a country is the percentage of people changing residence, that is, moving from one household or dwelling to another. If equation (2) is verified, this leads to a simple relationship between the parameter k, the logarithm of the average number of households, H_n, per zone for a particular partitioning of space, n^2, the total number of households, H, and the probability of a move in this partitioned space. As we know that

$$n^2 = \frac{H}{H_n}, \text{ we can write}$$

$$CMI_{H_n} = k \ln\left(\frac{H}{H_n}\right) = k\left(\ln H - \ln H_n\right) \tag{6}$$

where the average number of households per zone i is calculated as

$$H_n = \sum_{i=1,n^2} \frac{H_i}{n^2} \tag{7}$$

Now, when each household occupies a separate zone, H_n is equal to 1. In that case the probability of a residential move will be

$$CMI_H = k \ln\left(H\right) \tag{8}$$

so that

$$k = \frac{CMI_H}{\ln\left(H\right)} \tag{9}$$

From this equation we can define k as the probability of a residential move divided by the logarithm of the total number of households, and more generally from equation (6), as the probability of a move between zones divided by the logarithm of the number of zones, since

$$k = \frac{CMI_{H_n}}{\ln(n^2)}$$

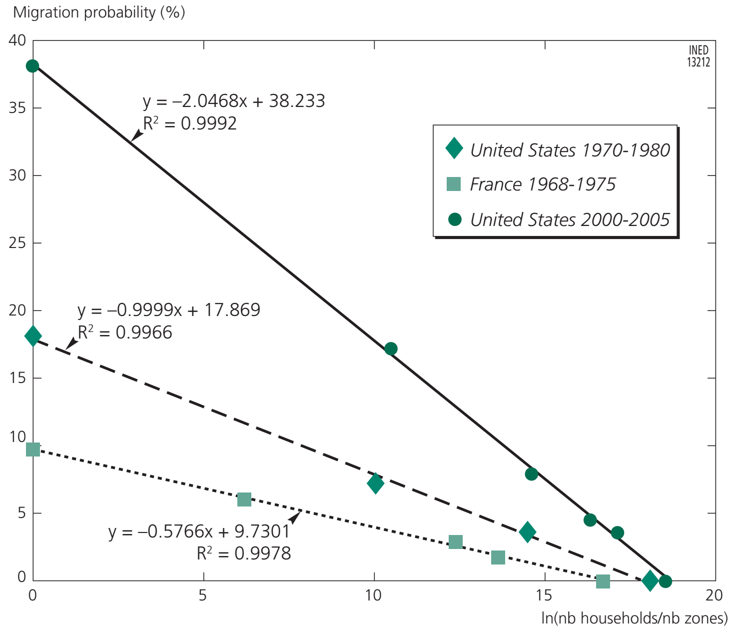
It follows that when we have no direct measure of the overall probability of an individual changing their place of residence, we may estimate this aggregate crude migration indicator (which we designate here as CMI_H) using k, provided we also have a count of the total number of households. Will it still be possible to derive a reliable indication of this overall migration intensity if the assumptions underpinning this simple relationship are not verified?

Some insight into this question can be derived by comparing the results obtained from directly measuring this index, where such data are available, against its value estimated by equation (6). We can first use the results obtained in Courgeau (1982), taking a different perspective on the data, to see if this point is verified.

Figure 1 shows the relationship between the instantaneous mobility rates, derived from the migrant-migration model, at different levels of spatial disaggregation, and the logarithm of the average number of households in each area, for 1970-1980 in the United States, and for 1968-1975 in France. The regression lines and their equations are also given, with the corresponding R^2 values. In the case of France we have four observations for the CMI corresponding to migration between regions (n = 22), departments (n = 95), communes (n = 36,394) and households (16,814,000), whereas for the United States the zonal systems are states (51), counties (3,084) and households (71,120,000). Note that the key innovation in this approach is that the units on the x-axis are now expressed in terms of the average number of households per zone (logged), in place of the number of zones (logged), as in Courgeau (1973b) and in Bell and Muhidin (2011). As a result, the y-intercept now represents the implied crude aggregate migration intensity CMI_H, since when each household occupies its own individual zone, household/zones = 1 and ln(household/zones) = 0.

The linearity of the relationship is strongly verified with values of R^2 greater than 0.995 for both countries. We can then recalculate the equations, this time excluding individual mobility, to estimate CMI_H, as though this figure were not available. For France we obtain a value of 9.73, which is very close to the observed value of 9.74, while for the United States the calculated value is 17.87, compared with an observed value of 18.30. We may hypothesize that the superior results for France are at least in part a product of the larger number of data points (4 zonal systems) than are available for the United States (3 zonal systems), and this interpretation is further supported by estimates for the period 2000-2005, where equation (6), computed over 5 data points, delivers an estimated CMI_H of 38.23, compared with the observed value of 38.08.

Figure 1. Migration probabilities by zonal system, for the United States (instantaneous for the period 1970-1980, and five-year for the period 2000-2005) and France (instantaneous for the period 1968-1975)



Source: Courgeau (1982), Current Population Survey (2005).

The use of annual mobility rates leads to a slightly lower estimate of the overall probability of changing residence, but has the advantage of being less distorted by multiple or return moves in such a short period. Application of the same approach to five-year or lifetime mobility will generate estimates of aggregate mobility in a similar fashion, but they are more dependent on multiple and return moves and therefore less reliable for international comparison. As indicated earlier, comparability depends on the parameters of the migrant-migration equation, which may be very different from one country to another.

We now have to examine what occurs when relationship (2) is not verified, and to establish whether a mix of the estimated parameters can be defined to retrieve the probability of a residential move. When a variation in density is introduced, it becomes necessary to use two parameters in model (3). This may also occur for other reasons, as found by Bell and Muhidin (2011):

> Other things being equal, positive intercepts imply a greater tendency towards long-distance migration than might otherwise be expected (since they are driven by higher intensities over smaller levels of disaggregation), whereas negative intercepts suggest that migration over long distances is less prevalent than expected, given the level of mobility over shorter distances.

In this case, the previous relationship (6) becomes:

$$CMI_H = k\ln(H)+b \tag{10}$$

and again allows straightforward estimation of the probability of a residential move when an independent measure of overall residential mobility is not available from the census.

We can illustrate the implementation of this model using data for three countries (Cambodia, Canada and Portugal) for which we have one-year migration data from IPUMS at three or more levels of spatial disaggregation, as needed to estimate two parameters,(8) and two countries with population registers (Belgium and the Netherlands: see Courgeau et al., 1989). Table 2 sets out estimates of the CMI_H values and the R^2 values of the linear regression. For Cambodia and Portugal there are no observed values against which to assess these estimates. However, the Canadian census collects information on all moves, delivering a figure of 13.43%, which is substantially higher than the estimate of 10.15% derived from equation (7). This underestimate by equation (7) may be due to the use of an exponent equal to 2 in the Pareto model, less strongly supported by the Canadian data. In contrast, for Belgium and the Netherlands, the estimates are very close to the observed values.

Table 2. Estimation of the CMI_H index (%) and the R^2 values for the linear regression

Country	Number of zones	Estimated CMI_H	Measured CMI_H	R^2
Cambodia (1998)	3	10.42	–	0.951
Canada (2001)	3	10.15	13.43	0.981
Portugal (2001)	4	8.51	–	0.945
Belgium (1983)	3	10.01	10.04	0.950
Netherlands (1983)	4	10.04	10.59	0.996

Source: Courgeau et al. (1989), Bell and Muhidin (2011).

If, as in the case of Canada, the log-relationship is not linear, this may call for an exponent other than 2 in the Pareto migration law. In this case, provided that data are available for a sufficiently large number of zoning systems, it is possible to estimate the parameters $k''' = k''f(\alpha)$ and α given in equation (4). The previous relationship then becomes:

$$\ln\left|1-\frac{CMI_H}{k'''}\right| = \left|2-\alpha\right|\ln(H) \tag{11}$$

and again allows an estimate of the probability of a residential move. Such an estimation is more complex than elaborated above, but as there are always two parameters to estimate, it has the same precision.

(8) For Canada and Portugal, the data are collected as migration transitions, while for Cambodia the estimates use the question on duration of residence.

Conclusion

We have been able to show that while the parameter k alone may not represent a clear, plain-language index of internal migration, combining k with other parameters, estimated from linear or nonlinear regression, generates a measure of the overall probability of changing residence which is internationally comparable. However such comparisons need to be made over a short period of time (instantaneous if using a migrant-migration model or otherwise using transition data for a single year). Multi-year or lifetime measures of migration inevitably introduce additional parameters that compromise international comparability. In the case of a multi-year transition period it is necessary to estimate the parameters of a migrant-migration model which will vary according to the country under observation. For lifetime migration, there is no straightforward solution because migration intensity accrues over individual lifetimes and is therefore dependent upon population age composition as well as on the ever-changing propensity for return and repeat migrations.

We have also argued that to provide a sound estimate of the probability of changing residence using this approach we need a large number of observations, corresponding to multiple levels of zonation. Two zonings will not permit a reliable estimate of this probability when two parameters need to be estimated, and a larger number of zonal systems is to be preferred. It is also useful to note that migrations are measured on the basis of administrative zones which, having a social reality, obey stronger regularities than random zones. Finally, while it is always possible to create different zonings based on existing areas, it is not possible to create zonings smaller than the one corresponding to the smallest area, for example in France, communal zoning.

The approach proposed in this paper needs to be tested further to confirm the validity of the model and the precision of the resulting estimates. One avenue worthy of exploration is to use models other than the Pareto, such as an exponential or a logarithmic-normal distribution (Kulldorff, 1956). Ultimately, however, it seems likely that an index capturing the overall probability of the propensity to change residence is the only reliable way to compare internal migration between countries. Its development therefore merits further scrutiny in the pursuit of a more rigorous understanding as to why levels of mobility differ between countries.

Appendices

Appendix 1. The main hypotheses and results for the theoretical spatial model used

We can first examine what occurs when we have a square territory with one side equal to a, with a total population, P, uniformly distributed and a total area, $S = a^2$: the population density will be equal to

$$\delta = \frac{P}{S} = \frac{P}{a^2}.$$

The whole territory is divided into n^2 zones.

Under these theoretical conditions, we also have to consider that migrations which do not cross the boundaries of a zone will not be registered, and that the frontiers of the whole country impose constraints which depend on the point of departure.

Let us first estimate the probability that:

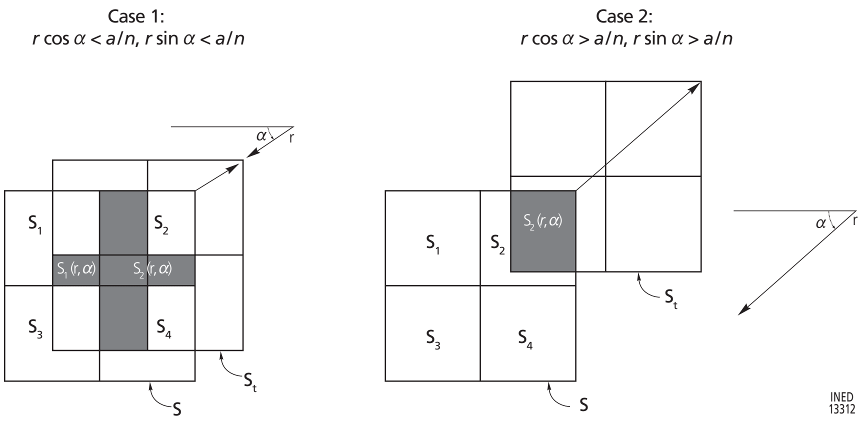
1) two points P_1, P_2 being randomly chosen in the country are in different areas;

2) the distance between them lies in the interval $(r, r + dr)$;

3) the angle between a given vector Ox and P_1P_2 will lie in the interval $(\alpha, \alpha + d\alpha)$.

For $n = 2$, Figure A.1.1 gives two cases, out of four different cases, of areas from which a migration may be recoded (first constraint) and is possible (second constraint) obtained from the translation $(r, -\alpha)$ in a polar coordinate system.

Appendix figure A.1.1: Territory divided into four zones with the areas from which a migration may be recorded, by migration distance



Source: Courgeau (1973b)

For the areas in this diagram, the probability to be estimated is:

$$p=\sum_{i=1}^{n^2}\frac{S_i}{S}\times\frac{S-S_i}{S}\times\frac{S_i\left(r,\alpha\right)}{S_i}\times\frac{r\,dr\,d\alpha}{S-S_i}=\frac{r\,dr\,d\alpha}{S^2}\times\sum_{i=1}^{n^2}S_i\left(r,\alpha\right)\qquad(1)$$

The value of the function $S_i\ (r,\alpha)$ differs in each case.

When $r\cos\alpha<a/n$ **and** $r\sin\alpha<a/n$ we have to introduce two constraints, so that

$$S\left(r,\alpha\right)=\sum_{i=1}^{n^2}S_i\left(r,\alpha\right)=\left(a-r\cos\alpha\right)\left(a-r\sin\alpha\right)-n^2\left(\frac{a}{n}-r\cos\alpha\right)\left(\frac{a}{n}-r\sin\alpha\right)$$
$$=a\left(n-1\right)r\left(\cos\alpha+\sin\alpha\right)-\left(n^2-1\right)r^2\cos\alpha\sin\alpha\qquad(2)$$

When $r\cos\alpha\geq a/n$ **or** $r\sin\alpha\geq a/n$ the second constraint no longer applies so that

$$S_0\left(r,\alpha\right)=\left(a-r\cos\alpha\right)\left(a-r\sin\alpha\right)\qquad(3)$$

Since the law of migration is independent of the direction of the move, integrating equation (1) according to α delivers the probability that:

a) two points randomly chosen in S will lie in two different areas, and

b) the distance between these two points will lie between r and $r+dr$.

We now have to distinguish different cases according to the value of r.

When $0\leq r<a/n$ the two constraints apply irrespective of the direction of the move, so that this integration leads to

$$f_1\left(r\right)dr=\frac{8r\,dr}{S^2}\int_0^{\pi/4}S\left(r,\alpha\right)d\alpha=\frac{dr}{S^2}\left[8a\left(n-1\right)r^2-2\left(n^2-1\right)r^3\right]\qquad(4)$$

When $a/n\leq r<a\sqrt{2}/n$ the angle $\varphi=\text{Arccos}\,a/nr$ distinguishes between the case when both constraints apply $\left(\varphi\leq\alpha\leq\pi/4\right)$ and the case when only the constraint imposed by the territory boundaries will apply $\left(0\leq\alpha<\varphi\right)$ so that

$$f_2\left(r\right)dr=\frac{8r\,dr}{S^2}\left[\int_0^{\varphi}S_0\left(r,\alpha\right)d\alpha+\int_{\varphi}^{\pi/4}S\left(r,\alpha\right)d\alpha\right]=\frac{dr}{S^2}\Big[8\,r\,a^2\,\text{Arccos}\frac{a}{nr}-8\,a\,r^2$$
$$+2\,r^3\left(n^2+1\right)-8\,a\,n\,r^2\sqrt{1-a^2/n^2r^2}+4a^2r\Big]\qquad(5)$$

When $a\sqrt{2}/n\leq r<a$, only the constraint imposed by the territory boundaries will apply, so that

$$f_3\left(r\right)dr=\frac{8r\,dr}{S^2}\int_0^{\pi/4}S_0\left(r,\alpha\right)d\alpha=\frac{dr}{S^2}\left(2a^2\pi\,r-8a\,r^2+2\,r^3\right)\qquad(6)$$

Finally, when $a\leq r\leq a\sqrt{2}$, no internal migration may occur, so that

$$f_4\!\left(r\right)dr=\frac{8r\,dr}{S^2}\int_{\varphi}^{\pi/4}S_0\!\left(r,\alpha\right)d\alpha=\frac{dr}{S^2}\left[2a^2r\left(\pi-2-4\ \text{Arccos}\frac{a}{r}\right)+8ar\sqrt{r^2-a^2}-2r^3\right]\qquad(7)$$

Now introducing the law of migration $m(r)=\dfrac{K}{r^2}$, we obtain the number of recorded migrations, $M(n^2)$:

$$M(n^2)=\left[\int_{r=0}^{a/n}\frac{K}{r^2}f_1(r)\,dr+\int_{r=a/n}^{a\sqrt{2}/n}\frac{K}{r^2}f_2(r)\,dr+\int_{r=a\sqrt{2}/r}^{a}\frac{K}{r^2}f_3(r)\,dr+\int_{a}^{a\sqrt{2}}\frac{K}{r^2}f_4(r)\,dr\right] \quad (8)$$

The last integral is independent of n; the other three functions are easy to integrate with the exception of

$$\frac{1}{r}\,\mathrm{Arccos}\,\frac{a}{nr}$$

which appears on the second integral, but substituting the variable $s=1/nr$ this becomes:

$$\int_{1/a}^{1/a\sqrt{2}}-\frac{1}{s}\,\mathrm{Arccos}(as)\,ds \quad (9)$$

which is independent of n.

After integration, this leads to the very simple equation:

$$M(n^2)=\frac{KP^2}{S^2}\left(C+2\pi a^2\ln(n)\right) \quad (10)$$

When n = 1 and we are considering the whole country, then the number of internal migrants is zero. We deduce that C = 0 and equation (10) becomes

$$CMI(n^2)=\frac{M(n^2)}{P}=K\pi\delta\ln(n^2) \quad (11)$$

When we consider a triangular territory, a similar calculation leads to the same equation, which suggests that the relationship (11) holds when a territory is divided into areas of the same shape and size.

Let us now always consider a square territory, but this time divided in two rectangular areas of different densities, δ_1 and δ_2. A similar calculation leads to a slightly different equation:

$$CMI(n^2)=\frac{M(n^2)}{P}=K\pi(\delta_1+\delta_2)\left[\frac{\delta_1^2+\delta_2^2}{(\delta_1+\delta_2)^2}\ln(n^2)+K'\frac{(\delta_1-\delta_2)^2}{(\delta_1+\delta_2)^2}\right] \quad (12)$$

In this case the number of square areas considered will be a multiple of 4, and the equation will no longer hold for n = 1. We can also easily verify that when $\delta_1 = \delta_2 = \delta$, we again derive equation (11).

If we now introduce a value for α different from 2 but retain square zones with constant density, and replacing $\dfrac{K}{r^2}$ by $\dfrac{K}{r^\alpha}$ in equation (8) after integration, this becomes:

$$CMI\left(n^2\right)=\frac{M\left(n^2\right)}{P}=\frac{KP}{a^{\alpha}}\left(1-\frac{1}{n^{2-\alpha}}\right)f\left(\alpha\right) \tag{13}$$

where $f(\alpha)$ is a function of α, independent from K, a and n. However when α remains close to 2.0 we can write:

$$\left(1-\frac{1}{n^{2-\alpha}}\right)=\ln\left(n^2\right)+g\left(n,\alpha\right) \tag{14}$$

where $g\left(n,\alpha\right)$ is infinitesimally small as α tends to 2, and the equation (13) will again closely approximate equation (11).

Appendix 2. The main hypotheses and results for the theoretical temporal model

Let us examine the hypotheses that lie behind the migrant-migration model (Courgeau, 1973a) which is derived from the mover-stayer model (Blumen et al., 1955).

First, the basic hypothesis of the mover-stayer model is that only a fraction, K, of a population that has made a previous migration is at risk of making a subsequent move. Its specific application to a group which had already made a migration, leads to the probability of making a new move, μ, independent of the order of that move. Such a hypothesis has been widely verified. Other assumptions of the migrant-migration model suppose that the instantaneous migration rate, m, in the total population, P, is time-invariant during the period of observation, and that return migrations form a constant proportion, l, of the migrations of order greater than one. These assumptions are very often verified for short periods of time.

Under these assumptions, over a very short interval $(\theta, \theta + d\theta)$ this population will make $P\ m\ d\theta$ migrations. Under the mover-stayer model, only a proportion of these movers will go on to make a new migration: $P\ m\ K\ d\theta$. Let us consider the distribution over time of these additional moves. During the time interval $(t, t + dt)$ these migrations will satisfy the following equation:

$$\frac{d\left[M_n\left(t\right)\right]}{P\,m\,K\,d\theta-M_n\left(t\right)}=\mu\,dt \tag{1}$$

where $M_n(t)$ represents the new migrations occurring between θ and t. Integrating this equation between θ and t we obtain this number of migrations, then by varying θ between an initial point in time (0) and a final point in time (t) we obtain an estimate of total new migrations occurring in this period:

$$\int_{\theta=0}^{\theta=t}P\,m\,K\left[1-\exp\left(-\mu\left\{t-\theta\right\}\right)\right]d\theta=P\,m\,K\left[t-\frac{1}{\mu}\left(1-\exp\left\{-\mu\,t\right\}\right)\right] \tag{2}$$

Assuming there is no return migration, then by calculating the difference between the total migrations recorded during this period ($M(t) = P\, m\, t$) and these multiple migrations, we obtain a first estimate of the number of migrants during the period:

$$m'(t) = P\,m\left[(1-K)t + \frac{K}{\mu}\left(1-\exp\{-\mu t\}\right)\right] \tag{3}$$

and then introducing return migrations as a constant proportion, l, of the migrations of order greater than 1, we obtain a final estimate of the number of migrants:

$$m(t) = P\,m\left[\left(1-K\{1+l\}\right)t + \frac{K(1+l)}{\mu}\left(1-\exp\{-\mu t\}\right)\right] \tag{4}$$

In order to estimate the parameters of this model we need data from detailed migration surveys, as in France (Courgeau, 2006), or from censuses measuring the number of migrants during different time periods, as in the United States (Long and Boertlein, 1990). With these parameters we can compare migration across a number of countries.

REFERENCES

BACCAÏNI Brigitte, COURGEAU Daniel, DESPLANQUES Guy, 1993, "Les migrations internes en France de 1982 à 1990 : comparaison avec les périodes antérieures", *Population*, 48(6), pp. 1771-1790.

BELL Martin, MUHIDIN Salut, 2009, "Cross-national comparisons of internal migration", United Nations, New York, Human Development Research Paper 2009/30, 66 p., http://hdr.undp.org/en/reports/global/hdr2009/papers/HDRP_2009_30.pdf

BELL Martin, MUHIDIN Salut, 2011, "Comparing internal migration between countries using Courgeau's *k*", in Stillwell J., Clarke M. (eds.), *Population Dynamics and Projection Methods*, Dordrecht, Heidelberg, London, New York, Springer, pp. 141-164.

BELL Martin, CHARLES-EDWARDS Elin, KUPISZEWSKI Marek, STILLWELL John, ZHU Y., 2011, "Internal migration around the world: Towards a global inventory", Plenary paper to the *6th International Conference on Population Geographies*, Umeå University, Sweden.

BELL Martin, BLAKE Marcus, BOYLE Paul, DUKE-WILLIAMS Oliver, REES Phil et al., 2002, "Cross-national comparison of internal migration: Issues and measures", *Journal of the Royal Statistical Society*, A, 165, pp. 435-464.

BLUMEN isadore, KOGAN Marvin, MCCARTHY Philip, 1955, *The Industrial Mobility of Labour as a Probability Process*, Ithaca, New York, Cornell University Press, 63 p.

COURGEAU Daniel, 1970, *Les champs migratoires en France*, Ined/PUF, Cahier 58, 158 p.

COURGEAU Daniel, 1973a, "Migrants et migrations", *Population*, 28(1), pp. 95-129.

COURGEAU Daniel, 1973b, "Migrations et découpages du territoire", *Population*, 28(3), pp. 511-537.

COURGEAU Daniel, 1979, "Migrants and migrations", selected papers, *Population*, 3, pp. 1-35.

COURGEAU Daniel, 1982, "Comparaison des migrations internes en France et aux États-Unis", *Population*, 37(6), pp. 1184-1188.

COURGEAU Daniel, 2006, "Mobility and spatial heterogeneity", in *Demography: Analysis and synthesis*, *Volume I*, Caselli, G., Vallin, J., Wunsch G. eds., Academic Press: Elsevier, pp. 279-292.

COURGEAU DANIEL, VAN DER ERF R. F., POULAIN Michel, 1989, "Mobilité et découpage géographique : comparaison entre Belgique et Pays-Bas", *L'Espace géographique*, 1, pp. 39-53.

CURRENT POPULATION SURVEY, 2005, "Table 1, General mobility by region, sex, age: 2000-2005", : http://www.census.gov/hhes/migration/data/cps.html

DONZEAU Nathalie, PAN KÉ SHON Jean-Louis, 2009, "Residential mobility trends in France, 1973-2006: new estimates", *Population-E*, 64 (4), pp. 687-704.

GINSBERG Ralph B., 1972, "Critique of probabilistic models: Application of the semi-Markov model to migration", *Journal of Mathematical Sociology*, 2, pp. 63-82.

HÄGERSTRAND Torsten, 1957, "Migration and area", in Hannerberg David, Hägerstrand Torsten, Odeving Bruno (eds.), *Migration in Sweden: A Symposium*, Lund Studies in Geography, Ser. B, 13, pp. 27-158.

HOLT David, STEEL David, TRANMER Mark, WRIGLEY Neil, 1996, "Aggregation and ecological effects in geographically based data", *Geographical Analysis*, 28, pp. 244-261.

JAKOBSON Arne, 1969, "*Omflyttingen i Sverige*, 1950-1960", Monographiserie, Lund, Sweden, Berlingska Boktryckeriet.

KITSUL Pavel, PHILIPOV Dimiter, 1981, "The one-year/five-year migration problem", in Rogers A. (ed.), *Advances in Multiregional Demography*, Laxenburg, Austria, International Institute for Applied Systems Analysis, pp. 1-33.

KULLDORFF Gunnar, 1955, *Migration Probabilities*, Lund, Sweden, Lund Studies in Geography, 14, C.W.K. Gleerup Publishers.

LEE Everett, 1966, "A theory of migration", *Demography*, 3(1), pp. 47-57.

LONG John F., BOERTLEIN Cecilia G., 1990, "Comparing migration measures having different intervals", Washington DC, US Bureau of the Census, *Current Population Reports*, Series P-23, Special Studies 166, pp. 1-11.

MYERS George C., MCGINNIS Robert, MASNICK George, 1967, "The duration of residence approach to a dynamic stochastic model of internal migration: A test of the axiom of cumulative inertia", *Eugenics Quarterly*, 4(2), pp. 121-126.

NOWOK Beata, 2010, *Harmonization by Simulation: A Contribution to Comparable International Migration Statistics in Europe*, Amsterdam, Rozenberg http://dissertations.ub.rug.nl/faculties/rw/2010/b.nowok/

NOWOK Beata, WILLEKENS Frans, 2011, "A probabilistic framework for harmonisation of migration statistics", *Population, Space and Place*, 17(5), pp. 521-533.

OPENSHAW Stan, 1984, *The Modifiable Areal Unit Problem,* Norwich, Geo Books, 40 p.

RAVENSTEIN Ernst G., 1885, "The laws of migration", *Journal of the Royal Statistical Society*, 48, pp. 167-235.

RAVENSTEIN Ernst G., 1889, "The laws of migration", *Journal of the Royal Statistical Society*, 52, pp. 241-305.

REES Philip, KUPISZEWSKI Marek, 1999, *Internal Migration and Regional Population Dynamics in Europe: A Synthesis*, Strasbourg, Council of Europe Publishing, 113 p.

ROGERS Andrei, RAYMER James, NEWBOLD K. Bruce, 2003, "Reconciling and translating migration data collected over time intervals of differing widths", *Annals of Regional Science*, 37(4), pp. 581-601.

ROGERSON Peter A., 1990, "Migration analysis using data with time intervals of differing widths", *Papers of the Regional Science Association*, 68, pp. 97-106.

ROYER Jean-François, 2009, "Estimating repeat and return migrations among sub-populations in France", *Population-E*, 64 (4), pp. 705-728.

STOUFFER Samuel A., 1940, "Intervening opportunities: A theory relating mobility and distance", *American Sociological Review*, 5, pp. 845-867.

ZIPF George K., 1946, "The P_1P_2/D hypothesis: On the intercity movement of persons", *American Sociological Review*, 11, pp. 677-686.

Daniel Courgeau, Salut Muhidin, Martin Bell • Estimating Changes of Residence for Cross-National Comparison

This short paper considers a number of temporal and spatial models that can be used for international comparison of internal migration levels across all countries of the world. First, among the various spatial models used, the model linking migrations to the zoning of the territory provides a simple summary of this relationship, but its parameters do not have a clear plain language meaning for international comparison. Second, the "migrant-migration" model derives an instantaneous rate based on migrant numbers measured over variable durations that is independent of multiple and return moves occurring over a longer interval. International comparison is thus only possible for the instantaneous mobility rate (change of residence), a standard indicator whose meaning is clear. The authors use numerous examples to show that the simultaneous use of both types of models provides a means, under certain conditions, to approximate such a rate, that can be linked to the parameters of these models. The validity of these models can be tested and confirmed using data from countries where direct measures of changes of residence are available.

Daniel Courgeau, Salut Muhidin, Martin Bell • Estimer les changements de résidence pour permettre les comparaisons internationales

Cet article envisage divers modèles, tant spatiaux que temporels, afin de permettre une comparaison internationale des niveaux de migration interne entre tous les pays du monde. En premier lieu, parmi les divers modèles spatiaux utilisés, le modèle reliant les migrations aux découpages du territoire permet de résumer simplement cette relation. Mais ses paramètres n'ont pas de signification claire pour une comparaison internationale. En second lieu, le modèle « migrant-migration » permet de ramener des effectifs de migrants, mesurés sur des périodes variables, à un taux instantané indépendant des migrations multiples et des retours survenant sur une période plus large. Dès lors, une comparaison internationale n'est possible que pour les taux instantanés de changement de logement, indicateur classique dont la signification est claire. Les auteurs montrent à l'aide de nombreux exemples que l'utilisation simultanée des deux types de modèles permet, sous certaines conditions, d'estimer de façon approchée un tel taux, que l'on peut relier aux paramètres de ces modèles. Dans certains pays, où l'on dispose d'une mesure directe des changements de logement, la validité des modèles peut être testée et confirmée.

Daniel Courgeau, Salut Muhidin, Martin Bell • Estimar los cambios de residencia para permitir las comparaciones internacionales

Este artículo considera diversos modelos, tanto espaciales como temporales, a fin de obtener una comparación internacional de los niveles de migración interna, entre todos los países del mundo. En primer lugar, entre los diversos modelos espaciales utilizados, el modelo que asocia las migraciones a las divisiones territoriales permite resumir simplemente esta relación. Pero sus parámetros no tienen una significación clara en una comparación internacional. En segundo lugar, el modelo "migrante-migración" permite reducir los efectivos de migrantes, medidos sobre periodos variables, a una tasa instantánea independiente de las migraciones múltiples y de los retornos acaecidos durante un periodo más amplio. Así pues, una comparación internacional solo es posible sobre las tasas instantáneas de cambio de residencia, indicador clásico de clara significación. Con la ayuda de numerosos ejemplos, los autores muestran que la utilización simultánea de los dos tipos de modelo permite, bajo ciertas condiciones, estimar dicho tipo de tasa, que se puede asociar entonces a los parámetros de los modelos utilizados. En ciertos países, en los que se dispone de una medida directa de los cambios de residencia, la validez de los modelos puede ser testada y confirmada.

Keywords: International comparaison, internal migration, spatial model, temporal model, change of residence, migrant-migration model, zoning.