On distance and the spatial dimension in the definition of internal migration

Niedomysl, Thomas; Fransson, Urban

Published in:
Annals of the Association of American Geographers

DOI:
10.1080/00045608.2013.875809

2014

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

This manuscript is a post-print (ie final draft post-refereeing) that has been published. Please cite: Niedomysl, T. & Fransson, U. (2014) On distance and the spatial dimension in the definition of internal migration. Annals of the Association of American Geographers 104 (2) 357-372.

On Distance and the Spatial Dimension in the Definition of Internal Migration

Abstract

Migration is commonly defined by a temporal and a spatial dimension. It is generally agreed that these dimensions are problematic and may seriously distort understanding of migration, but data constraints have effectively obstructed further insights. This article focuses on the spatial dimension where migration is typically defined as movement across administrative borders. Borders usually serve as proxies for migration distance, but the validity of such proxies is largely unknown, posing a considerable challenge to migration research. Using data for all internal migrants in Sweden, the only known country where migration distances are available in sufficient detail, we present the first accurate description of actual migration distances and investigate the relationship between actual migration distances and migration-defining boundaries. More specifically, we examine how a) the volumes of migration and b) the characteristics of migrants change when migration distances vary and when different types of migration-defining boundaries are employed. The findings show that notable shares of short-distance migrants are included almost regardless of which migration-defining boundary is employed, but migrant differentials are less affected than might be expected. Key Words: *migration; distance; administrative boundaries; migrant differentials*

Introduction

The study of human migration is a topic of significant interest across the social sciences. This is hardly surprising considering that migration is arguably the most important form of capital mobility and affects the lives of those involved in a most profound way. While various disciplines tend to specialize in certain areas, they share a common understanding that migration involves two dimensions: time and space. The temporal dimension relates to how long someone has to stay at a new location to be regarded as a migrant, and the spatial dimension relates to how far someone has to move to be counted as a migrant. Migration scholars also share an understanding that each dimension is problematic and potentially constitutes serious obstacles for achieving a better understanding of migration. The problems are not merely related to determining "how long" or "how far"; in fact, they are more fundamental than that. Data constraints have effectively prevented a thorough investigation of these two dimensions, precluding further insights and with the result that the actual scope and consequences of these problems are largely unknown.

This article focuses on the spatial dimension in the definition of internal migration. The key issue here is distance, which is arguably critical for understanding the volumes, causes, and effects of migration for individual migrants and societies. For obvious reasons it matters how far someone moves, as those who move long distances are more likely to experience significant change and consequences compared to those who move only short distances. It is also known that what causes someone to move across a short-distance is very different from what causes a long-distance move where housing concerns dominate the former type of movement and mainly employment the latter (Morrison and Clark, 2011; Niedomysl, 2011). The problem is, however, that data on migration distances are not available from standard censuses. Individuals are conventionally defined as migrants when moving across some administrative border, regardless of the distance of the move. Borders thus serve as proxies for migration distance, but are also of functional relevance in the sense that e.g. tax rates and public services usually vary across administrative areas. To add further complexity, administrative areas vary considerably between and within countries in terms of size, shape, and settlement pattern, not to mention that adjacent areas also vary in these regards, making the validity of borders as proxies for distance most uncertain. Such uncertainty, and the risk of confusing short-distance migrants with long-distance migrants, poses a considerable challenge for improving understanding of migration, in particular when considering that these two groups may have little else in common than having changed their address.

The aim of this article is to investigate the differences between migrants who move over certain distances and migrants who move between different administrative areas at different spatial scales. Using longitudinal micro data for all internal migrants in Sweden, the only known country where sufficiently detailed information on migration distances and migrant characteristics is available, we examine how a) the volumes of migration and b) the characteristics of migrants change when migration distances vary as well as when different types of migration-defining boundaries are employed.

Using data where migration distances are measured with unprecedented accuracy, this research will allow us to, first, determine the risk of confusing short-distance migrants with long-distance migrants under varying types of migration-defining boundaries and, second, to investigate the impact of using proxies for distance instead of actual distance on migrant differentials. Both aspects are fundamental for determining the validity of the spatial dimension in conventional definitions of migration.

Theory and Previous Research

Migration and distance has been of scholarly interest ever since Ravenstein (1885), in one of the earliest examples of the scientific study of internal migration, noted that most migrants proceed only a short distance. He thereby confirmed the opinion of Smith (1776 [1981]) that people constitute the most difficult form of capital to be transported over distance. Tellingly, Ravenstein also pointed out one of the key problems of migration research, in that migrants are defined as migrants when they cross administrative borders, but due to the varying size, shape, and situation of the areas, any analysis is fraught with difficulties:

"[A] circumstance likely to lead to misconception, if not error, arises from the very unequal size of the counties. Rutland and Yorkshire are hardly comparable. A journey of 25 miles at the most converts any native of Rutland into a "migrant," whilst a native of Yorkshire to place himself into the same position might have to travel as many as 95 miles." (Ravenstein, 1885:168)

Since then a number of studies have shown that there is a strong deterrent of distance to migration flows (see e.g., Makower et al., 1938; 1939; Hägerstrand, 1957; Olsson, 1965; Long et al., 1988a; 1988b), and because migration is still almost exclusively defined by administrative borders, the problems identified over a century ago are equally valid today. Nonetheless, migration researchers have had to try to cope with these problems, and distance has come to occupy a central place in migration research.

Distance also plays a central role in the development of more formal models of migration. Numerous studies have followed the early works of researchers such as Stouffer (1940), Zipf (1946), and Stewart (1948), in which gravity models were used to explain aggregate migration flows between regions (see e.g. Plane, 1984; Stillwell and Congdon, 1991). In brief, distance is treated as a proxy for variables that are difficult to measure, and by including other variables that are expected to be of importance for migration decision-making, estimates can be made regarding their relative impact on attracting or repelling migrants (for some recent examples see Congdon, 2010; Cooke and Boyle, 2011; Biagi et al., 2011; Kalogirou, 2012). While spatial interaction modeling of migration flows incorporate a measure of distance, it tends to be a very coarse measure, usually given little analytical weight. Exceptions do exist, however, and the work of Gordon (1991) is a good example where distinctive migration flows are separated and investigated to provide further insights. Nonetheless, insights gained from such exercises likely depend upon the accuracy of data on migration distance and the extent to which individual characteristics can be linked to migration flows, or more generally, on the level of data aggregation.

Central to this literature is the so called modifiable areal unit problem (MAUP) that arises with the aggregation of data into areal units (Openshaw, 1984; see Wong, 2009, for a recent overview). Two issues of concern are related to this problem. The scale effect refers to differences in results that may arise simply due to variation in the number of spatial units employed. The zonation effect refers to differences that arise as a result of the numerous ways in which those spatial units may be divided. Since migration data are usually only available at aggregate levels for administrative areas that vary in terms of size, shape, settlement pattern and situation, the MAUP applies and conclusions drawn based on such aggregate migration data may be very different from conclusions drawn based on data where the exact origins and destinations of individual migrants is available (Rogerson, 1990).[1] A related concept is that of ecological fallacy (Robinson, 1950), where fallacies may arise when inferences from results of groups are deduced to individuals, a problem that occurs "…because areal studies cannot distinguish between spatial associations created by the aggregation of data and real associations possessed by the individual data prior to spatial aggregation" (Openshaw, 1984:14). The MAUP and ecological fallacy may in fact be more serious in migration studies than in other fields since migration may take place across several administrative areas (as already discussed, distance is a key factor distinguishing migration from residential mobility).

[1] Another, partly related, issue pertains to changes in statistical boundaries which make it problematic to compare migration over time (see e.g. Blake, et al., 2000; Amcoff, 2006).

Theoretically, the finding that most people move only short distances has mainly been explained with reference to a) the monetary costs of moving longer distances, b) the fact that the amount of information on the potential new location decreases with increasing distance, and c) the psychic costs of moving (such as the loss of contacts with family and friends) (Greenwood, 1975). The notion that monetary costs would constitute a main obstacle, however, was questioned early on and it has been shown that direct transportation costs are negligible, even more so considering the potential economic gains of moving (Lansing and Mueller, 1967; Sahota, 1968).

Consequently, more credence has been given to lack of information and psychic costs in explaining the distance decay of migration (Ritchey, 1976). In an attempt to empirically estimate the relative importance of these two factors, Schwartz (1973) argued that education would offset the dearth of information (the ability to obtain and process information is assumed to increase with education) and that the psychic costs of long-distance migration would increase with age (older individuals have more invested in relations with family and friends). This predicts that if lack of information is the main deterrent, education would be positively related to migration distance, whereas age would have the opposite relation to distance if psychic costs were the main deterrent. Finding the expected relation between education and distance, but not between age and distance, Schwartz (1973:1153) suggested that "the adverse effect of distance on migration is basically a diminishing-information phenomena."

A quite different approach was suggested by Stouffer (1940, 1960), who introduced the concept of intervening opportunities, which proposed that the number of persons going a given distance is directly proportional to the percentage increase in opportunities at that distance. A similar train of thought is found in Zipf (1949), who argued the importance of the principle of least effort (see also Smith, 1978). Zipf's simple explanation is that most people will not go a longer distance than they have to in order to achieve their goals. This raises a number of questions as to which goals can be achieved over varying migration distances and whether any differences exist across population subgroups.

Although much attention has been paid to examine the characteristics of migrants compared to non-migrants (see Greenwood and Hunt, 2003, for an overview of the early literature) less work appears to have focused on how the characteristics change over distance. Rose (1958), however, constitutes one of the first examples. Categorizing migrants by socioeconomic status, he found that high status was positively correlated with migration distance, suggesting that individuals from high-status groups must move longer distances to achieve their goals (see also Bogue and Thompson (1949) for an earlier example). In a more recent study Hunt (2004) suggested that highly educated people are more likely to move longer distances compared to those with low levels of education.

Considering the above, it is not surprising to find variations in self-expressed motives for migration across distance. The work of Gleave and Cordey-Hayes (1977) constitutes one of the first examples of a study where information from survey data on self-expressed motives for migration were used to examine motivational variation across migration distance. In a review and empirical extension of this literature, Niedomysl (2011) found considerable variation in migration motives, not only over migration distance, but also particularly in relation to migrants' socioeconomic and demographic characteristics. The generalized results showed, however, that housing-related motives were negatively related to increasing migration distance, whereas the opposite was true for employment-related motives (see also Morrison and Clark, 2011).

As briefly noted earlier, migration literature is largely made up of studies examining migration flows between administrative areas where distance is typically included as an independent variable. However, findings are often inconclusive, sometimes showing insignificant and theoretically implausible results. Arguably, this could very well be a result of using administrative borders as proxies for distance, due to the risk of confusing short-distance migrants with long-distance migrants. Support for such a claim is found in the few studies that have made efforts to distinguish between short-distance and long-distance migrations, as these studies tend to come up with more believable results (Weeden, 1973; Gordon, 1975; DaVanzo, 1976a; 1976b; Stillwell, 1978; Jun and Chang, 1986; Gordon, 1991; Nivalainen, 2004; Biagi et al., 2011). For example, Jun and Chang (1986) made a distinction between migration flows between contiguous and non-contiguous administrative areas, to sort out short-distance from long-distance migration in the absence of data on actual distance (see also White and Meuser, 1988). They concluded that "noncontiguous migration behaves significantly different from contiguous migration, and that noncontiguous migration, rather than total migration, should be used for the study of factors affecting interstate migration" (Jun and Chang, 1986:17).

The work of White and Meuser (1988) is another example that emphasizes the implications of selecting administrative borders for investigating migration. They argue that the migration-defining boundary is fundamental in that it distinguishes between different types of social processes and is selective of population characteristics. Their results also show that different kinds of moves are highly dependent upon personal characteristics, in concurrence with human capital theories and life cycle theories of local residential mobility.

Keeping these results in mind, it is nonetheless striking how *little* impact these studies, where distance is taken into more than passing consideration, have had upon general migration research. A possible, but not a good, explanation for this neglect is that the reviewed studies have been obstructed by a lack of high quality data on migration distances. With few exceptions they have been forced to use the administrative borders approach since individuals' migration distances have not been available. In addition, they have relied on data from relatively small samples of migrants. It is possible that this has restricted their impact on migration research. In summing up this brief review, it is evident that even if migration distance appears to be quite vaguely understood from a theoretical perspective, it plays a crucial role for the fundamental ways migration is analyzed, and by extension, for all conclusions drawn as well. Regardless of how distance should be understood from a theoretical perspective, for the individual migrant, distance obviously matters when it is no longer possible to see on a regular basis those people and places that used to be part of everyday life (Roseman, 1971). For progress to be made, more in-depth data on migration distances have to be employed.

By employing uniquely detailed migration data, this article will address three specific research questions to provide more accurate descriptions of the importance of the spatial dimension in the definition of migration: 1. What does the distribution of migrants across actual migration distance look like? 2. How does that distribution change when different types of (administrative) migration defining borders are employed? 3. What do the individual characteristics of migrants across actual migration distance, and across different types of (administrative) migration defining borders, look like?

Data

Identifying Migrants and Inferring Moving Distances

In Sweden, all individuals have according to law to be registered as living in unique real property units. The centroids of all properties have known exact coordinates collected by the Swedish mapping, cadastral and land registration authority. For confidentiality reasons, the coordinates of the properties are truncated into 100 by 100 meter squares when transferred to Statistics Sweden and made available for research. This means that the exact coordinate is not known to researchers (only the location of a property within a 100 by 100 meter coordinate square). It is the centroids of the coordinate squares that constitute the starting point for estimations of moving distances in this article and the whole of Sweden is covered by this 100 by 100 meter square grid. Since Sweden is a sparsely populated country, a vast majority of squares do not contain any properties at all whereas in the more populated parts of Sweden, a square can contain up to 57 properties (though this is an extreme outlier). In addition, some large properties cover areas larger than the 100 by 100 meter square grid.

In 2008, a total of 1 241 519 individuals changed their 100 by 100 meter square of residence (the square of residence are known at the end of each year, i.e. the 31st of December, in the data at our disposal). When limiting the population to those 18 years or older, 1 000 363 individuals had changed their square of residence. Since the whole of Sweden is covered by the 100 by 100 meter square grid, Euclidian moving distances between the squares can be calculated with considerable accuracy: the minimum moving distance to end up in an adjacent square is 1 meter and the maximum distance if moving to an adjacent square is 282 meters. But since the exact coordinates of the properties are not available, distances are inferred between the centroids of the origin and destination squares. Nonetheless, all moving distances can be inferred with considerable accuracy (the maximum error is only ±142 meters) and at this level of precision the modifiable areal unit problem is not a concern.[2]

A more detailed presentation with illustrations of inferred Euclidian moving distances between six different properties is provided in Figure 1 and Table 1. For simplicity, we assume that each migration takes place over a distance of at least 1 meter and that all properties are located within one square (that they do not extend across the borders of the 100 by 100 meter squares). Hence, the distance between the centroids for a migrant moving from

[2] For readers who are used to working with origin-destination migration matrixes, instead of the data employed here, it may be instructive to know that these data potentially provide a migration matrix with more than 3.6 *billion* inhabited cells.

his/her origin to destination D_1 is 100 meters (inferred distance in Table 1). However, since the exact location of the properties O and D_1 are not known, the actual moving distance can range from 1 meter to 224 meters (if the exact location of the origin and destination is very close or very distant to the border).[3] Similarly, if moving from O to D_2, the distance between the centroids is 141 meters, but could in theory be 1 meter or 282 meters. Furthermore, Table 1 shows that the maximum error (regardless of actual migration distance) does not exceed 142 meters.

(FIGURE 1 ABOUT HERE)

(TABLE 1 ABOUT HERE)

However, a problem may arise if properties are extended (e.g. if someone buys a neighboring property and the latter property obtains the same property code as the original property). In such cases, the coordinates of the property may change so that a property located in one 100 by 100 meter square in 2007 may appear in an adjacent square in 2008. Hence, even if no one actually moved, the coordinates may change, making it falsely appear as if someone had moved. Conversely, the same problem may arise if a property is parceled out (e.g. if someone sells part of their property and the new property obtains a different property code).

To minimize the risk that no such “false” movers were included in the data two restrictions were made. First, we set the minimum inferred moving distance to be greater than 100 meters, thus blocking out adjacent 100 by 100 meter squares as shown in Figure 2. This restriction reduced the number of movers to 858 584 individuals and sets the minimum inferred moving distance to 141 meters (see Table 1), though it is still possible to move 1

[3] Note that we use the term “actual migration distance” simply to denote Euclidian distance measured with considerable accuracy. It should not be taken to imply that we know anything about the routes of transportation.

meter and end up in a destination square located diagonally from the square of origin. Second, for those remaining individuals who had the same property code in 2008 as in 2007, but had changed their 100 by 100 meter square between the two years, we added a further restriction that they would have to move a distance greater than 1000 meters. This further reduced the number of movers to 846 320. These measures may perhaps seem overly cautious, but we wanted to be sure that no “false” movers were included in the data, even if it meant reducing the population of movers from 1 000 363 to 846 320. However, these measures also remove an unknown number of “true” movers. It may be noted that some “true” movers are not possible to detect in the data we have access to. This latter is due to the fact that some properties cover more than one 100 by 100 meter square and may contain many residents. In such cases, an individual may indeed move, but still be registered as a resident at the same property code, thus not appearing as a mover in the data set. To the authors best knowledge, no data exists that would allow us to estimate how many “true” movers go undetected, but they presumably constitute a negligible share.

(FIGURE 2 ABOUT HERE)

To identify movers across administrative borders (discussed later), layers of the administrative units are simply merged onto the 100 by 100 meter square grid, so that each individual square is identifiable within specific administrative units. Movers can then be identified from the square of their origin, across any administrative border, to the square of their destination. It should also be noted that although Euclidian migration distances can be measured with a maximum error margin of ±142 meters, for simplicity, all reported distances in the article are rounded to the nearest kilometer.

Although the Swedish register data (discussed further in the next section) are uniquely detailed and of exceptionally high quality for migration research purposes, they are not without flaws (Amcoff, 2009). In addition to the risk of including “false” movers discussed earlier, two general issues need to be mentioned here. First, in the data at our disposal, only one migration event is recorded per year and the exact date of the event is not available. Second, individuals are required by law to report a “permanent” change of residence to the authorities, but some individuals are likely to fail to do so (mainly those who have not established themselves in the housing market). Both issues lead to an undercount of migration volumes (presumably short-distance migration), but the extent of these problems and hence their relevance for analyses of migration distance is not known.

In this article we make a distinction between short-distance (residential mobility) and long-distance migration based on whether a migrant changes his/her location of workplace when moving (see Zax, 1994). Since there is no exact way to determine a distance at which a combined change of residence and workplace is “optimal”, we examined the proportion of employed movers that changed the location of their workplace by distance of move. The results (available from the authors) showed that the increase in the proportion that changed their workplace location leveled off at an approximate moving distance of 100 km and we thus define migration as a move exceeding 100 km. This is a definition that also resonates with general notions on the distinction between short-distance and long-distance migration, discussed earlier, since a move exceeding 100 km is most likely to involve a significant change in the daily life of the individual concerned (Hägerstrand, 1957; Roseman, 1971).

The Individual Characteristics of Migrants

Information on the individual characteristics of migrants is drawn from the Swedish register data provided to researchers by Statistics Sweden. The term “register data” refers to data collected on individuals by various Swedish government agencies for administrative purposes and updated on a continuous basis (and some of this data are transferred to Statistics Sweden). In Sweden, all legal inhabitants obtain a unique social security code from birth or, for immigrants, when a permanent residence permit is granted. Using each individuals' social security number as a unique key, Statistics Sweden can link the various pieces of data concerning each legal inhabitant of Sweden over time. Researchers who seek access to this data must specify their demands and Statistics Sweden will, after a confidentiality assessment, provide a database where the individuals' social security numbers have been replaced by randomly generated unique id-codes. This provides researchers with access to a wealth of data on socioeconomic and demographic characteristics for all individuals (including the location of their properties described in the previous section). For a detailed overview of the Swedish register data, see Statistics Sweden (2011).

The register data we employ in this article covers the entire adult Swedish population in 2008. Table 2 below shows the individual characteristics we selected and these are standard characteristics reported in the migration literature. Table 2 shows both migrants and non-migrants.

Three demographic variables, *sex*, *age*, and whether *children* are present in the household, were included. Sex and children are self-explanatory, but it may be worth mentioning that the four age classes were based on a rough estimate of different life phases that may influence migratory behavior.

Moreover, four socioeconomic variables were included. The level of *education* was divided into four classes. While compulsory school in Sweden currently encompasses nine years, older people may have two additional years of schooling, but this is no longer the case. Upper-secondary school (twelve years) is now the standard and this explains why twelve years constitute a class of its own, before it is possible to embark upon a university education.

Those with university degrees were divided into those with degrees completed in less than four years, and those with degrees completed in four or more years.

To measure *income,* the Swedish population’s disposable incomes (the amount left after taxes and transfers) were divided into quartiles (the first quartile had less than 99,000 SEK and the fourth quartile had more than 200,000 SEK in disposable income per year).[4] Information on whether the migrant *changed job* in connection with migration was also included, as was information on *occupation*. The latter was divided into four groups: employed (including self-employed), students, unemployed, and an “other” category that mainly included retirees and those on long-term sick leave. It should be noted that, unlike the other characteristics, change of job and occupation are measured on the 1st of November in the data available. This means that there is an important undercount (two months) of those who change jobs in particular. Finally, included in Table 1 is the key factor of interest, *migration distance,* which was divided into nine categories of incrementally increasing length.

(TABLE 2 ABOUT HERE)

The differences between migrants and non-migrants in Table 2 are as expected from the literature on migrant differentials. Clearly, these two groups are different in many regards as only gender and presence of children in the household have an equal balance. As regards the other characteristics, in comparison to non-migrants, migrants tend to be much younger with 34 per cent in the age group 18-25 years, compared to 10 per cent in the corresponding age group among non-migrants. Some differences, although not as clear, are noted as regards education where there is a slight tendency for migrants to have a higher level of education.

[4] 1 SEK = 0.15 USD (17 June 2013)

Further, migrants are considerably more inclined to have changed jobs during the last ten months, more likely to have been unemployed or student prior to moving and to have lower incomes (the latter in part due to their younger age and in part due to the greater risk of having been unemployed).

Administrative Regions

Six different types of administrative regions are employed in the forthcoming analyses (number of units in brackets): parish (1785), municipality (290), LA region (78), county (21), NUTS 2 (8), and NUTS 1 (3). The *parish* is an administrative level mainly used by the Swedish church (the Swedish church is nowadays separated from the Swedish state), but many municipalities still use it for analytical purposes (with negligible exceptions, parishes merge into municipalities) and e.g. Statistics Sweden presents annual reports on the number of inhabitants for all parishes. *Municipalities* constitute the most important administrative level in Sweden as municipalities are independent bodies with an elected local government, taxation rights and various responsibilities such as running childcare, schools and health services. Further, *LA regions* (local labor market regions) are based on commuting patterns and constructed from municipalities but do not merge into counties and have no administrative functions (see Amcoff (2009) for a critical discussion of LA regions). *Counties* constitute the second type of administrative regions that, along with municipalities, has real administrative functions – partly via the “Landsting” which overlaps counties – such as taxation rights and some responsibilities for public services (municipalities are merged into counties). The *NUTS 2* and *NUTS 1* regions (Nomenclature des Unités Territoriales Statistiques) are determined by the European Union for comparisons across member countries but are rarely used for internal analyses in Sweden (counties are merged into NUTS regions).

As far as it is known, NUTS regions have never been used for studies of migration in Sweden and they do not have any functional purpose.

Two maps of Sweden, showing the municipality and the NUTS 2 levels as examples, along with a table that presents statistics on area and population for the administrative levels are provided in Figure 3 and Table 3 below.

(FIGURE 3 ABOUT HERE)

(TABLE 3 ABOUT HERE)

Method

To answer the first two research questions, i.e. what does the distribution of migrants across actual migration distance look like and how that distribution changes when different administrative borders are employed as migration-defining boundaries, descriptive measures will be used. In addressing the third research question, i.e. how the characteristics of migrants vary across distance depending on whether actual distance or proxies for distance (administrative borders) are employed, we will use regression analysis.

More specifically, we use the following logistic regression model to calculate the odds ratio for an individual with certain characteristics to move across varying types of administrative borders/distance: odds(Y = 1) = $e^{\ln[\text{odds}(Y=1)]} = e^{(\alpha+\beta_1 X_1 + \beta_2 X_2 + \dots \beta_k X_k)}$. Where Y is the dependent variable (moving across a specific migration distance or a specific administrative border), $X_1, \dots, X_k$ constitute the independent variables (migrant characteristics) and $\beta_1, \dots, \beta_k$, are the regression parameters that determine the relationship between the odds ratio and the independent variables. Logistic regression is an appropriate method to use when the key dependent variables are categorical, as is the case with the distance categories and administrative regions employed in this article. When reporting the results of the regressions we present the relative odds, exp (β), i.e. the exponentiation of the β coefficient, and we also report the Nagelkerke R-square which is a “pseudo” R-square that can be used for comparisons across regressions (higher values suggest a greater fit of the model). Furthermore, for all regressions we report the number of migrants, also including those in the reference group (the numbers may seem large but this is because most migrants fall into the reference category). The shortest migration distance and smallest administrative areas are only used as reference categories in the regressions.

It may be noted that unlike actual distance, administrative borders, as proxies for distance, cannot be divided into comparable classes to fit a multinomial model (a migrant may cross different types of borders in a single move). Therefore, binary logistic regressions were employed to investigate how the characteristics of migrants vary across, first, actual distance (nine classes/models are used, fewer classes were also tested but the division of distance into nine classes presented the most distinct results) and, second, administrative borders (one regression for each type of administrative border, using movers within parish borders as the base for all regressions).

This approach, where a migrant may cross additional borders in a spatial administrative hierarchy, means that moves over the longest distances will be present in all administrative border regressions. Only shares of the shortest distances are gradually excluded when the spatial resolution decreases. Even though this somewhat disguises the effect of distance, the approach mimics conventional use of administrative borders as proxies for distance where, regardless of which administrative border is employed as a migration-defining boundary, the longest distances are always present. As far as the authors are aware, it is not possible to make a direct statistical test in one statistical model between actual migration distance and administrative borders as proxies for distance, for the simple reason that a large variation of actual migration distances are present in the administrative border regressions. Hence, only comparisons between the models are feasible.

To test the robustness of the regressions all independent variables were included in a stepwise manner and changes in the coefficients were analyzed. This resulted in a reduction of the independent variables and classes originally included. Moreover, a five per cent sample was tested, but the results were clearly unsatisfactory. This is quite expected considering the highly skewed distribution of migration distances whereby too few migrants moving longer distances are included when a small random sample is used. In countries where only samples of migration data are available, it is thus clearly advisable to use a sample of migrants stratified by migration distance. Regressions were also run where the island of Gotland (the largest island located off the south-east coast; see Figure 3 above) was excluded since anyone migrating from the island will have to move a distance of at least 100 km to reach the mainland (Gotland is a municipality, a local labor market region, and a county even though it holds only approximately 60,000 inhabitants). However, excluding Gotland did not have an effect on the outcome of the models and so migrants to and from Gotland were included.

Results

Descriptive Results

Table 4 below presents the first detailed description of migration distances for a whole country in which all movers moving a distance longer than 141 meters are included (first column) (columns 2-7 include only those who cross the administrative border as stated in the column heading). In 2008, more than 840,000 people changed their square of residence distributed over 190,000 areal units (i.e. 100 by 100-meter squares). The mean distance is 53 km, but the median is only five km, obviously a highly uneven dispersion. The fact that the mode is zero may perhaps seem confusing, but it reveals that the most common move is in fact shorter than 500 meters (all distances were rounded to the nearest kilometer). It is further notable that the minimum distance is zero (again, due to rounding) regardless of administrative border employed even when Sweden is divided into only three regions (column 7). Evidently this means that someone moved a distance shorter than 500 meters but happened to cross a NUTS 1 region border (the largest administrative area), showing that what may appear as merely a theoretical problem (i.e. including short-distance movers when employing large administrative areas) also exists in reality.

(TABLE 4 ABOUT HERE)

The number of migrants is reduced by almost 300,000 when parish borders are used as the migration-defining boundaries and the median distance increases notably to 14 km. When the municipality border is used as the migration-defining boundary – as is occasionally done in Swedish migration studies, in part because of data availability and in part because of the importance of municipalities as administrative units – distance increases further to a median distance of 55 km. As illustrated more clearly in Figure 4, a large share of migrants across municipality borders are not even close to being considered "long-distance" movers. Evidently, researchers should be wary of using municipality borders as migration-defining boundaries and these results clearly cast some doubt on studies that have done so unaware of the large share of short-distance movers.

Fortunately, as most policy implications drawn from research on internal migration in Sweden implicitly assume a focus on long-distance migration, it has become more common to use local labor market regions (LA region) as migration-defining boundaries. As shown in Figure 4 there is a notable difference in migration distance compared with municipalities where the median distance is 190 km for LA regions compared to only 55 km for municipalities. Even so, notably large shares of those who cross LA regional borders do not move very far. Approximately 20 per cent travel distances less than 81 km, which is arguably within acceptable commuting range in the most populous parts of Sweden with well-developed infrastructure.

The number of administrative areal units decreases from 78 to 21 when county borders are employed instead of LA regions, but migration distances increase only modestly and there is no longer a drastic reduction in the number of migrants. Now, the 20 per cent of migrants moving the shortest distance reach 92 km, which is likely to be a relatively cumbersome commuting distance and is close to the definition of migration as a move exceeding 100 km employed in the present article.

(FIGURE 4 ABOUT HERE)

The division of Sweden into 8 regions (NUTS 2) further increases migration distances, consolidating the mean and median values. It is, however, not known whether this regional division has ever been used in Swedish migration research, nor the last and broadest division (NUTS 1), which includes only three regions. Even so, for migration researchers concerned with not including a significant number of short-distance migrants, it appears as if the NUTS 1 regional division is the most feasible choice (at least if the ambition is to exclude short-distance migrants). The ten per cent of migrants that cross this type of border and move the shortest distances reach 141 km, and even the five per cent are in the vicinity of 100 km.

However, while the obvious expectations, that the fewer the number of administrative units, the fewer the movers and the longer the average distance they travel, are true, the picture is evidently more complicated. For example, considering the reduction in the number of administrative units when comparing LA regions (78 units) with Counties (21 units), it might be expected that substantial differences in the number of movers and their moving distances would appear (as was the case when comparing Municipalities (290 units) with LA regions (78 units)). However, only marginal differences arise, somewhat calling into doubt simplistic expectations about the relationship between number of administrative units and moving distances.

These descriptions clearly demonstrate that the administrative borders approach for defining migration is indeed problematic. A large share of “migrants” travel only across short distances, a pattern present almost regardless of border type. Hence, there is an obvious risk that migration scholars will confuse long-distance migrants with short-distance ones. While such evidence has not been available throughout the history of migration until now, it is far from evident what the implications of these findings are.

Regression Results

Table 5 presents the results of the logistic regressions for distance where movers across 0-9 km constitute the reference category in all models. The results show that women are somewhat less likely to move across longer distances than men. The greatest difference in this regard is seen for migration over 160-229 km, where women are 1.11 (1/0.90) times less likely to move. Recalling Table 2, where descriptive characteristics of migrants and non-migrants was presented, it is interesting to note that women actually showed a higher propensity of moving versus staying than men.

It is well established that young people constitute the most frequent migrant group (Rogers and Castro, 1981), and this pattern is evident also across migration distances. As a rule, age differences become more pronounced over longer distances, mainly longer than 110 km, but it is notable that the age group 41-64 years has about the same probability of moving over short distances (less than 70 km) as the youngest age group.

Similarly, it is not surprising to find that people who have children living in their household have a lower propensity to move longer distances. This does not mean that families with children are immobile: in fact they are 1.1 times (1/0.91) more likely to move over 10-19 km compared to those without children, but this relationship changes as distance increases. Presumably, this is related to parents' reluctance to uproot their children, which is inevitably the case when moving longer distances (although it could be argued that for small children in particular, even moving a few kilometers is likely to be a significant change). It should be noted, however, that there appears to be a threshold effect of distance in that migration distances exceeding 40 km do not change the estimates to any noteworthy extent (with the exception of the longest distances).

Regarding education, the results are somewhat mixed even though the general pattern is clear: the higher the level of education, the higher the probability of moving over longer distances. For the shortest moves, however, the highly educated display a lower probability of moving. Looking only at those with a university degree requiring four or more years of study (i.e. those with 17 or more years of schooling), the probability of moving is 1.65 times higher than for those with only a compulsory education (over 40-69 km) and increases to 3.75 times over 330 km. The finding that this group would stand out so markedly over the longest distances is clearly interesting and could possibly be explained by the uneven geography of universities in Sweden, although further research would be required to establish this.

(TABLE 5 ABOUT HERE)

(TABLE 6 ABOUT HERE)

While the results for education are largely in line with the literature, the results for income are not. Conversely, the general pattern is instead that of a lower probability of moving with increasing income. For example, migrants who are in the highest income quartile are approximately 1.7 to 2 times less likely to move distances longer than 75 km compared to migrants in the lowest income quartile. However, it should be kept in mind that the income variable does not measure *change* of income.

As expected, people are more likely to change jobs the farther they move. It is also expected that people who were unemployed prior to moving have higher odds of moving longer distances compared to those who were employed prior to moving. In this case it is interesting to note that this effect does not increase with a linear progression for the longest distances. Students have somewhat higher odds compared to the unemployed.

Turning to Table 6, where administrative borders are employed as proxies for distance, the results appear at first to be very similar to those of Table 5. However, upon closer inspection it is evident that there is less dispersion for the estimates of Table 6 compared to Table 5. Even though this is to be expected since there are fewer models in Table 6, the lesser dispersion in Table 6 does not seem to be random. Overall, the results are in agreement with the estimates gained from actual distance, but it should be remembered that the comparability between the distance categories in Table 5 and the administrative borders categories in Table 6 are not easily comparable (the latter categories cover a wide range of distances).

Women have a lower propensity to move over long distance (about the same odds as in Table 5). Turning to age, Table 6 shows consistently somewhat lower odds, suggesting a larger age-effect across administrative borders. For example, comparing the odds of the oldest age group in Table 5 with the equivalent in Table 6, the odds range from 0.74 to 0.44 in Table 5 and from 0.51 to 0.35 in Table 6.

Moreover, in Table 6, not having children living at home increases the probability of moving by 1.3 times, regardless of the type of administrative border. When actual distances were measured in Table 5, the odds were notably lower, and in the case of moving 10-19 km, even negative.

For education, it is evident that the higher the level of education, the higher the probability of moving longer distances. For example, having spent 17 years or more in school increases the probability of moving across a NUTS 1 border by 4.4 times compared to those with only compulsory education. Recalling Table 5, the odds became positive only for those moving distances longer than 40 km. Regarding income, the results are very similar when comparing Table 6 with Table 5.

Turning to job change, it is interesting to note that those moving across a parish border are 1.4 times more likely to change jobs compared to those moving within parish borders. Here, it is important to keep in mind that the median migration distance is only 14 km (see Table 4) for those moving across parish borders, hardly a distance where job change is obvious. Similarly, the odds (1.7) of changing jobs when moving across a municipality border are higher than might be expected, considering that large shares of this group move very short actual distances. The overall impression is that the administrative borders approach shown in Table 6 produces higher odds for job change than might be expected given the relatively large shares of short-distance migrants present for all types of areal units.

Finally, only minor differences are detectable for the last variable, occupation, when the odds in Table 6 are compared to those in Table 5.

Concluding Discussion

It has been 128 years since Ravenstein (1885) first identified the key problems with the spatial aspect of conventional definitions of migration. Meanwhile migration scholars have had little choice but to hope that these problems are not too serious. This article has cast new light on these issues by investigating the relationship between actual migration distances and migration-defining boundaries. Using uniquely detailed data on migration distances, this article has provided the first large-scale analysis of how, first, the volumes of migration and, second, the characteristics of migrants change when migration distances vary as well as when different types of migration-defining boundaries are employed. Two main findings stand out.

First, it is determined that a significant share of migrants move only short distances, even when migration-defining boundaries of relatively large administrative regions are employed. This finding confirms long-held qualms and is clearly a cause for validity concerns. In short, there is an obvious risk of confusing short-distance migrants with long-distance ones, increasing the potential danger of drawing wrong conclusions about many aspects of migration. Even if the data employed in this article are unique, there are no obvious reasons to believe that the findings would not be generalizable to other developed countries (at least those with similar population geography, i.e. a high degree of urbanization and a relative concentration of population to coastal areas). The evidence presented in this article thus suggests that much previous migration research may be significantly hampered due to failure of accounting for actual migration distance. For example, if a study employs a random sample population and defines migrants as those who have changed the location of their dwelling, only a tiny fraction of those that changed their dwelling will be long-distance migrants. If the study population of migration researchers is not the intended one, the ability to draw valid conclusions may of course be severely damaged.

A counter-strategy occasionally employed in previous research to mitigate the risk of including short-distance migrants has been to exclude migrants moving to adjacent administrative areas (usually referred to in the literature as non-contiguous migration). However, this strategy is rarely employed, possibly because little is known about the migration population included/excluded by such measures. More common are approaches that have tried to improve distance estimates, typically by measuring the centroids of administrative regions, sometimes weighted by population (Boyle and Flowerdew, 1997), thus going beyond the standard approach of simply assuming that movers across certain administrative region types are long-distance migrants. Of note is also the work of Rogerson (1990) who used geometrical probability to estimate migration distances under various assumptions related to the MAUP. However, the extent to which such measures correspond with actual migration distances remains to be explored. An appropriate next step for the present research would therefore be to investigate the extent to which such measures are suitable for providing a more reasonable estimation of migrants. This is important because in the vast majority of countries, only migration flow data between relatively large administrative areas are available, making it virtually impossible to use actual distance as a migration-defining criterion (see Bell et al., 2002).

Another interesting finding was the non-linear relationship between the number of administrative units and the number of migrants. A reduction in the number of administrative units does not automatically correspond with an expected reduction of migrants. This finding implies that at a certain point the different region types pick up similar population processes in spite of their considerable areal differences.

Second, and perhaps somewhat surprising, the comparative analyses of how migrants' characteristics change across actual distance and administrative borders were more consistent than might be anticipated considering the above and previous research (Jun and Chang, 1986; White and Meuser, 1988; Biagi et al., 2011). To be sure, differences do exist between the estimates gained from the two approaches, but do not appear to be of such magnitude as to be cause for great concern. At least not if the migration-defining boundary is chosen with care, taking into consideration the findings presented here. Hence, one interpretation of these findings is that even though there are obvious validity issues with the spatial definition of migration, previous research results on individuals' migration differentials appear to have been little affected by the inclusion of short-distance migrants. However, an alternative interpretation based on the fact that all migrants of an entire country were used and that a five per cent sample failed to produce satisfactory results, suggests that sample data on migration unstratified by distance will be less easily interpreted.

In addition, the article has advanced the literature by showing that migrant differentials exist also across distance/administrative borders, not only when investigating differences between migrants and non-migrants. The generalized results showed that women, being older, having children or high income are characteristics that decrease the propensity to move across longer distances. Conversely, being highly educated, having changed jobs, being unemployed or student, increased the propensity to move across longer distances. Again, there is little to suggest that these findings would not be generalizable to other countries, although caution is required since migrant differentials are more likely to be influenced by country-specific characteristics such as differences in welfare systems.

In essence then, while this study has confirmed validity concerns held since the late 19th century, the question of whether these concerns are significant enough to produce invalid conclusions on migration is still open for debate and empirical investigations. For further research, important extensions include to incorporate distance in analyses of the Swedish migration system where the explicit role of types of regional settings and the directionality of moves are explored. Lastly, however, it needs to be emphasized that only differences in terms of migrant characteristics have been investigated here, not other types of migration determinants nor the effects on individual migrants or the societal consequences of migration. Possibly, the relatively large shares of short-distance migrants present, almost regardless of the administrative migration-defining boundary employed, will have more serious consequences in such cases.

References

Amcoff, J. 2006. Rural population growth in Sweden in the 1990s: unexpected reality or spatial–statistical chimera? Population, Space and Place, 12 (3): 171-185.

Amcoff, J. 2009. Rapid regional enlargement in Sweden: a phenomenon missing an explanation. Geografiska Annaler: Series B, Human Geography, 91 (3): 275–287.

Bell, M., Blake, M, Boyle, P., Duke-Williams, O., Rees, P., Stillwell, J., & Hugo, G. 2002. Cross-national comparison of internal migration: issues and measures. Journal of the Royal Statistical Society: Series A (Statistics in Society), 165 (3):435-464.

Biagi, B., Faggian, A., & McCann, P. 2011. Long and short distance migration in Italy: the role of economic, social and environmental characteristics. Spatial Economic Analysis, 6 (1): 111-131.

Blake, M., Bell, M., & Rees, P. 2000. Creating a temporally consistent spatial framework for the analysis of inter-regional migration in Australia. International Journal of Population Geography, 6 (2): 155-174.

Bogue, D.J., & Thompson, W.S. 1949. Migration and distance. American Sociological Review, 14 (2): 236-244.

Boyle, P.J., & Flowerdew, R. 1997. Improving distance estimates between areal units in migration models. Geographical Analysis, 29 (2): 93-107.

Congdon, P. 2010. Random-effects models for migration attractivity and retentivity: a Bayesian methodology. Journal of the Royal Statistical Society: Series A (Statistics in Society), 173 (4): 755-774.

Cooke, T,J., & Boyle, P. 2011. The migration of high school graduates to college. Educational

Evaluation and Policy Analysis, 32 (2): 202-213.

DaVanzo, J. 1976a. Differences between Return and Nonreturn Migration - Econometric-Analysis. International Migration Review 10 (1): 13-27.

DaVanzo, J. 1976b. Why families move: a model of the geographic mobility of married couples. Santa Monica: Rand.

Gleave, D., & Cordey-Hayes, M. 1977. Migration dynamics and labour market turnover. Progress in Planning 8 (1): 1-95.

Gordon, I. 1975. Employment and housing streams in British inter-regional migration. Scottish Journal of Political Economy 22 (2): 161-177.

Gordon, I. 1991. Multi-stream migration modeling. In J. Stillwell and P. Congdon. Eds. Migration models: macro and micro approaches. Pp. 73-91. London, UK: Belhaven Press.

Greenwood, M.J. 1975. Research on Internal Migration in the United States: A Survey. Journal of Economic Literature, 13 (2): 397-433.

Greenwood, M.J., & Hunt, G.L. 2003. The early history of migration research. International Regional Science Review, 26 (1): 3-37.

Hunt, J. 2004. Are migrants more skilled than non-migrants? Repeat, return, and same-employer migrants. Canadian Journal of Economics, 37 (4): 830-849.

Hägerstrand, T. 1957. Migration and Area. In D. Hannerberg, T. Hägerstrand & B. Odeving (Eds.), Migration in Sweden: a symposium (pp. 27-158). Lund: C.W.K. Gleerup Publishers.

Jun, I.S., & Chang, H.S. 1986. Functional forms and the relevance of contiguous migration in the study of migration and employment growth. The Annals of Regional Science, 20 (2): 17-27.

Kalogirou, S. 2012. Testing local versions of correlation coefficients. Jahrbuch für regionalwissenschaft, 32 (1): 45-61.

Lansing, J.B., & Mueller, E. 1967. The Geographic Mobility of Labor. Ann Arbor: University of Michigan, Institute for Social Research.

Long, L., Tucker, C.J., & Urton, W. L. 1988a. Measuring Migration Distances: Self-Reporting and Indirect Methods. Journal of the American Statistical Association, 83 (403): 674-678.

Long, L., Tucker, C.J., & Urton, W. L. 1988b. Migration Distances: An International Comparison. Demography, 25 (4): 633-640.

Makower, H., Marschak, J., & Robinson, H.W. 1938. Studies in Mobility of Labour: A Tentative Statistical Measure. Oxford Economic Papers, 1 (1): 83-123.

Makower, H., Marschak, J., & Robinson, H.W. 1939. Studies in Mobility of Labour: Analysis for Great Britain, Part I. Oxford Economic Papers, (2): 70-79.

Morrison, P.S., & Clark, W.A.V. 2011. Internal migration and employment: macro flows and micro motives. Environment and Planning A, 43 (8): 1948-1964.

Niedomysl, T. 2011. How migration motives change over migration distance: evidence on variations across socioeconomic and demographic groups. Regional Studies, 45 (6): 843-855.

Nivalainen, S. 2004. Determinants of family migration: short moves vs. long moves. Journal of Population Economics, 17 (1): 157-175.

Olsson, G. 1965. Distance and Human Interaction: A review and bibliography. Philadelphia, PA, Regional science research institute.

Openshaw, S. 1984. The Modifiable Areal Unit Problem. Norwich, Geobooks.

Plane, D.A. 1984. Migration Space: Doubly Constrained Gravity Model Mapping of Relative Interstate Separation. Annals of the Association of American Geographers, 74 (2): 244-256.

Ravenstein, E.G. 1885. The Laws of Migration. Journal of the Royal Statistical Society, 48 (2): 167-235.

Ritchey, N.P. 1976. Explanations of migration. Annual Review of Sociology, (2): 363-404.

Robinson, W.S. 1950. Ecological Correlations and the Behavior of Individuals. American Sociological Review, 15 (3): 351-357.

Rogers, A., & Castro, L.J. 1981. Model migration schedules. Laxenburg: International institute for applied systems analysis.

Rogerson, P.A. 1990. Buffon's needle and the estimation of migration distances, Mathematical Population Studies: An International Journal of Mathematical Demography, 2 (3): 229-238.

Rose, A. M. 1958. Distance of Migration and Socioeconomic Status of Migrants. American Sociological Review, 23 (4): 420-423.

Roseman, C.C. 1971. Migration as a spatial and temporal process. Annals of the Association of American Geographers, 61 (3): 589-598.

Sahota, G.S. 1968. An Economic Analysis of Internal Migration in Brazil. Journal of Political Economy, 76 (2): 218-245.

Schwartz, A. 1973. Interpreting the Effect of Distance on Migration. The Journal of Political Economy, 81 (5): 1153-1169.

Smith, A. 1776 [1981]. An Inquiry into the Nature and Causes of the Wealth of Nations. Chicago, Phoenix Books.

Smith, T.E. 1978. A cost-efficiency principle of spatial interaction behavior. Regional Science and Urban Economics, 8 (4): 313-337.

Statistics Sweden. 2011. Integrated database for labour market research. 1990–2009. Background Facts, Labour and Education Statistics 2011:4, Statistics Sweden.

Stewart, J.Q. 1948. Demographic Gravitation: Evidence and Applications. Sociometry, 11

(1/2): 31-58.

Stillwell, J.C.H. 1978. Interzonal Migration: Some Historical Tests of Spatial-Interaction Models. Environment and Planning A, 10 (10): 1187-1200.

Stillwell, J.C.H. & Congdon, P. 1991. Migration models: macro and micro approaches. London, UK: Belhaven Press.

Stouffer, S.A. 1940. Intervening Opportunities: A Theory Relating Mobility and Distance. American Sociological Review, 5 (6): 845-867.

Stouffer, S.A. 1960. Intervening opportunities and competing migrants. Journal of Regional Science, 2 (1): 1-26.

Weeden, R. 1973. Inter-regional migration models and their application to Great Britain. National Institute of Economic and Social Research, Paper 2. Cambridge, UK: Cambridge University Press.

White, M.J. & Mueser, P.R. 1988. Implications of Boundary Choice for the Measurement of Residential Mobility. Demography, 25 (3): 443-459.

Wong, D. 2009. The Modifiable Areal Unit Problem (MAUP). In A.S. Fotheringham & P.A. Rogerson (Eds.) The Sage Handbook of Spatial Analysis (pp. 105-124). Los Angeles, CA: Sage.

Zax, J.S. 1994. When is a move a migration? Regional Science and Urban Economics, 24 (3): 341-360.

Zipf, G.K. 1946. The PP/D hypothesis: on the intercity movement of persons. American Sociological Review, 11: 677-686.

Zipf, G.K. 1949. Human Behavior and the Principle of Least Effort. Addison-Wesley.

Table 1 Distance calculations of the examples given in Figure 1

Inferred distance (*between centroids*)	Min. distance	Max. distance	Max. error
O to D_1 = 100 meters	O to D_1 = 1 meter	O to D_1 = 224 meters	O to D_1 = 124 meters
O to D_2 = 141 meters	O to D_2 = 1 meter	O to D_2 = 282 meters	O to D_2 = 141 meters
O to D_3 = 200 meters	O to D_3 = 101 meters	O to D_3 = 316 meters	O to D_3 = 116 meters
O to D_4 = 282 meters	O to D_4 = 142 meters	O to D_4 = 424 meters	O to D_4 = 142 meters
O to D_5 = 223 meters	O to D_5 = 101 meters	O to D_5 = 360 meters	O to D_5 = 137 meters

Note: O = Origin, D = Destination.

Table 2 Descriptive characteristics

	Non-migrants		Migrants	
Category	N	%	N	%
Gender				
Male	3014239	49.2%	420128	49.6%
Female	3111884	50.8%	426192	50.4%
Age				
18-25	602762	9.8%	284124	33.6%
26-40	1439696	23.5%	305751	36.1%
41-64	2659424	43.4%	187691	22.2%
65+	1424241	23.2%	68754	8.1%
Children living at home				
Yes	2545456	41.6%	357599	42.3%
No	3580650	58.4%	488721	57.7%
Education				
11 years or less	3328507	54.3%	335379	39.6%
12 years	1079900	17.6%	242933	28.7%
13-16 years	1608858	26.3%	256541	30.3%
17 or more years	108857	1.8%	11467	1.4%
Income				
First quartile	1054866	17.2%	232796	27.5%
Second quartile	1394021	22.8%	192204	22.7%
Third quartile	1270829	20.7%	169225	20.0%
Fourth quartile	2406406	39.3%	252095	29.8%
Job change				
No change	4986149	81.4%	554559	65.5%
Job change	1139974	18.6%	291761	34.5%
Occupation				
Employed	3808784	62.2%	589989	69.7%
Unemployed	82468	1.3%	99741	11.8%
Student	207294	3.4%	80635	9.5%
Other	2027577	33.1%	75955	9.0%
Moving distance (km)				
0-9	-	-	514058	60.7%
10-19	-	-	98417	11.6%
20-39	-	-	62804	7.4%
40-69	-	-	35761	4.2%
70-109	-	-	24657	2.9%
110-159	-	-	20185	2.4%
160-229	-	-	24601	2.9%
230-329	-	-	22518	2.7%
330+	-	-	43319	5.1%
Total	6126123		846320	

Table 3 Descriptive statistics for the administrative areas, 2008

	Parish	Municipality	LA region	County	NUTS 2	NUTS 1
Number of areal units	1785	290	78	21	8	3
Area (Mean) km^2	n.a.	1522	5809	21024	55188	147169
Area (Median) km^2	n.a.	692	4201	11422	38953	80823
Area (Std. Dev.) km^2	n.a.	2668	5554	24639	50532	143186
Population (Mean)	5186	31918	121792	440771	1157024	3085412
Population (Median)	2074	15258	37111	273382	1096085	3526807
Population (Std. Dev.)	7928	62477	310926	499434	612935	1223094
Population density (Mean) km^2	n.a.	131	22	42	66	42
Population density (Median) km^2	n.a.	25	15	30	30	50
Population density (Std. Dev.) km^2	n.a.	458	29	62	96	34

Note: The total area of Sweden is 441506 km^2. No up-to-date maps at parish level were available.

Table 4 Migration distances in kilometers across different administrative borders, 2008

	All movers	Parish	Municipality	LA region	County	NUTS 2	NUTS 1
Number of areal units	190092	1785	290	78	21	8	3
Number of migrants	846320	555955	301453	156311	141050	119613	74233
Mean	53	80	141	250	268	297	380
Median	5	14	55	190	212	245	349
Mode	0	2	16	66	62	62	396
Std. Dev.	131	155	189	209	213	216	222
Min.	0	0	0	0	0	0	0
10 percentile	0	2	9	51	58	68	141
20 percentile	1	3	15	81	92	116	197
30 percentile	1	6	22	112	132	163	247
40 percentile	3	9	33	151	171	201	294
60 percentile	9	21	90	240	261	299	397
70 percentile	17	42	160	307	335	376	453
80 percentile	41	108	247	398	410	447	513
90 percentile	173	265	401	513	519	549	687
Max.	1498	1498	1498	1498	1498	1498	1498

Note: Except for the two first rows, all values are rounded to the nearest kilometer. For further details see text.

Table 5 Results of binary logistic regressions for migration across distance (migration across 0-9 km is reference category in all regressions)

	Migration distance							
	10-19 km	20-39 km	40-69 km	70-109 km	110-159 km	160-229 km	230-329 km	330+ km
Category	Exp(B) Sig.	Exp(B) Sig.	Exp(B) Sig.	Exp(B) Sig.	Exp(B) Sig.	Exp(B) Sig.	Exp(B) Sig.	Exp(B) Sig.
Sex (*ref. = male*)								
Female	0.99	0.99	0.96***	0.94***	0.92***	0.90***	0.94***	0.93***
Age (*ref. = 18-25*)								
26-40	0.95***	0.88***	0.75***	0.68***	0.63***	0.65***	0.64***	0.76***
41-64	1.01	1.06***	0.98	0.87***	0.69***	0.66***	0.65***	0.67***
65+	0.74***	0.62***	0.59***	0.50***	0.52***	0.48***	0.44***	0.45***
Children at home (*ref. = yes*)								
No	0.91***	1.03**	1.12***	1.11***	1.05**	1.08***	1.06***	1.17***
Education (*ref. = 11 years*)								
12 years	1.03**	1.02	1.08***	1.14***	1.32***	1.39***	1.50***	1.50***
13-16 years	0.93***	0.85***	1.22***	1.40***	1.62***	1.87***	1.93***	2.17***
>= 17 years	0.86***	0.84***	1.65***	2.09***	2.28***	2.99***	3.20***	3.75***
Income (*ref. = first quartile*)								
Second quartile	0.90***	0.90***	0.85***	0.82***	0.86***	0.84***	0.82***	0.79***
Third quartile	0.90***	0.90***	0.77***	0.71***	0.70***	0.67***	0.64***	0.62***
Fourth quartile	0.94***	0.95***	0.81***	0.65***	0.60***	0.59***	0.57***	0.58***
Job change (*ref. = no job change*)								
Job change	1.09***	1.21***	1.52***	1.87***	2.12***	2.32***	2.34***	2.56***
Occupation (*ref. = employed*)								
Unemployed	0.87***	0.96**	1.07***	1.15***	1.18***	1.10***	1.14***	1.14***
Student	0.80***	0.88***	1.24***	1.40***	1.54***	1.53***	1.52***	1.48***
Other	0.98	1.10**	1.31***	1.43***	1.35***	1.31***	1.52***	1.47***
Constant (B)	-1.465	-1.992	-2.703	-3.105	-3.366	-3.263	-3.395	-2.935
N	612475	576862	549819	538715	534243	538659	536576	557377
Nagelkerke R2	0.005	0.006	0.018	0.039	0.058	0.074	0.076	0.100

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table 6 Results of binary logistic regressions for migration across administrative borders (migrants within parish is reference category in all regressions)

	Migration across type of border					
	Parish	Municipality	LA region	County	NUTS 2	NUTS 1
Category	Exp(B) Sig.	Exp(B) Sig.	Exp(B) Sig.	Exp(B) Sig.	Exp(B) Sig.	Exp(B) Sig.
Sex *(ref. = male)*						
Female	0.96***	0.95***	0.92***	0.93***	0.93***	0.93***
Age *(ref. = 18-25)*						
26-40	0.83***	0.78***	0.66***	0.67***	0.68***	0.69***
41-64	0.77***	0.73***	0.65***	0.62***	0.61***	0.62***
65+	0.51***	0.40***	0.38***	0.36***	0.35***	0.37***
Children at home *(ref. = Yes)*						
No	1.32***	1.33***	1.36***	1.35***	1.37***	1.36***
Education *(ref. = 11 years)*						
12 years	1.18***	1.24***	1.35***	1.40***	1.42***	1.48***
13-16 years	1.65***	1.81***	2.01***	2.27***	2.37***	2.44***
>= 17 years	2.05***	2.44***	3.46***	3.84***	4.17***	4.44***
Income *(ref. = first quartile)*						
Second quartile	0.85***	0.79***	0.78***	0.77***	0.76***	0.75***
Third quartile	0.77***	0.69***	0.62***	0.60***	0.59***	0.58***
Fourth quartile	0.86***	0.80***	0.60***	0.62***	0.60***	0.58***
Job change *(ref. = no job change)*						
Job change	1.42***	1.74***	2.30***	2.34***	2.43***	2.58***
Occupation *(ref. = employed)*						
Unemployed	1.00	1.02**	1.18***	1.14***	1.15***	1.15***
Student	1.18***	1.29***	1.55***	1.57***	1.57***	1.57***
Other	1.02	1.11***	1.30***	1.34***	1.37***	1.40**
Constant (B)	0.507	-0.153	-0.901	-1.057	-1.265	-1.790
N	846320	591818	446676	431415	409978	364598
Nagelkerke R2	0.051	0.092	0.158	0.168	0.173	0.165

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.00$.

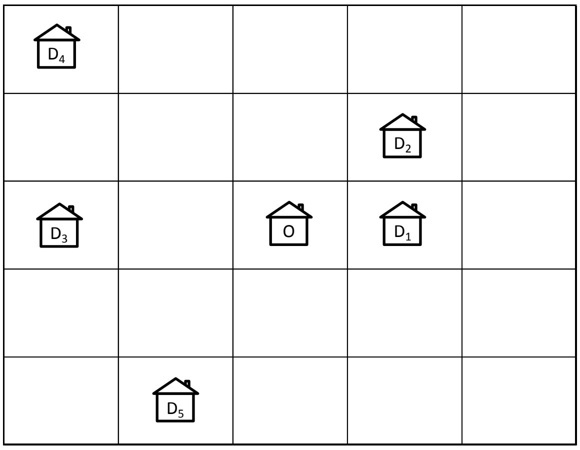

Figure 1 Illustration of coordinate squares used for calculating migration distances in Sweden
Note: O = Origin, D = Destination. Each square represents 100 by 100 meters.

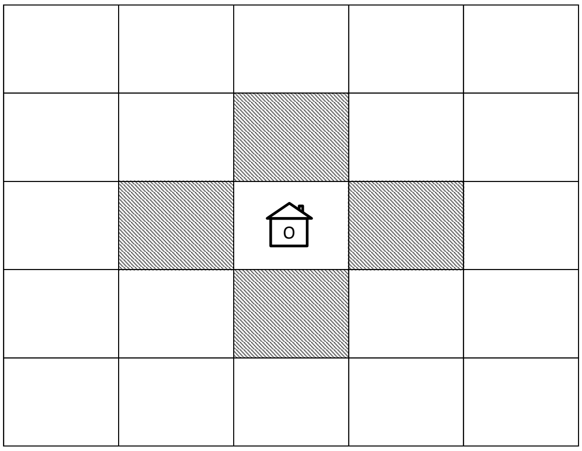

Figure 2 Setting a restriction that inferred distances have to be greater than 100 meters
Note: O = Origin. Each square represents 100 by 100 meters.

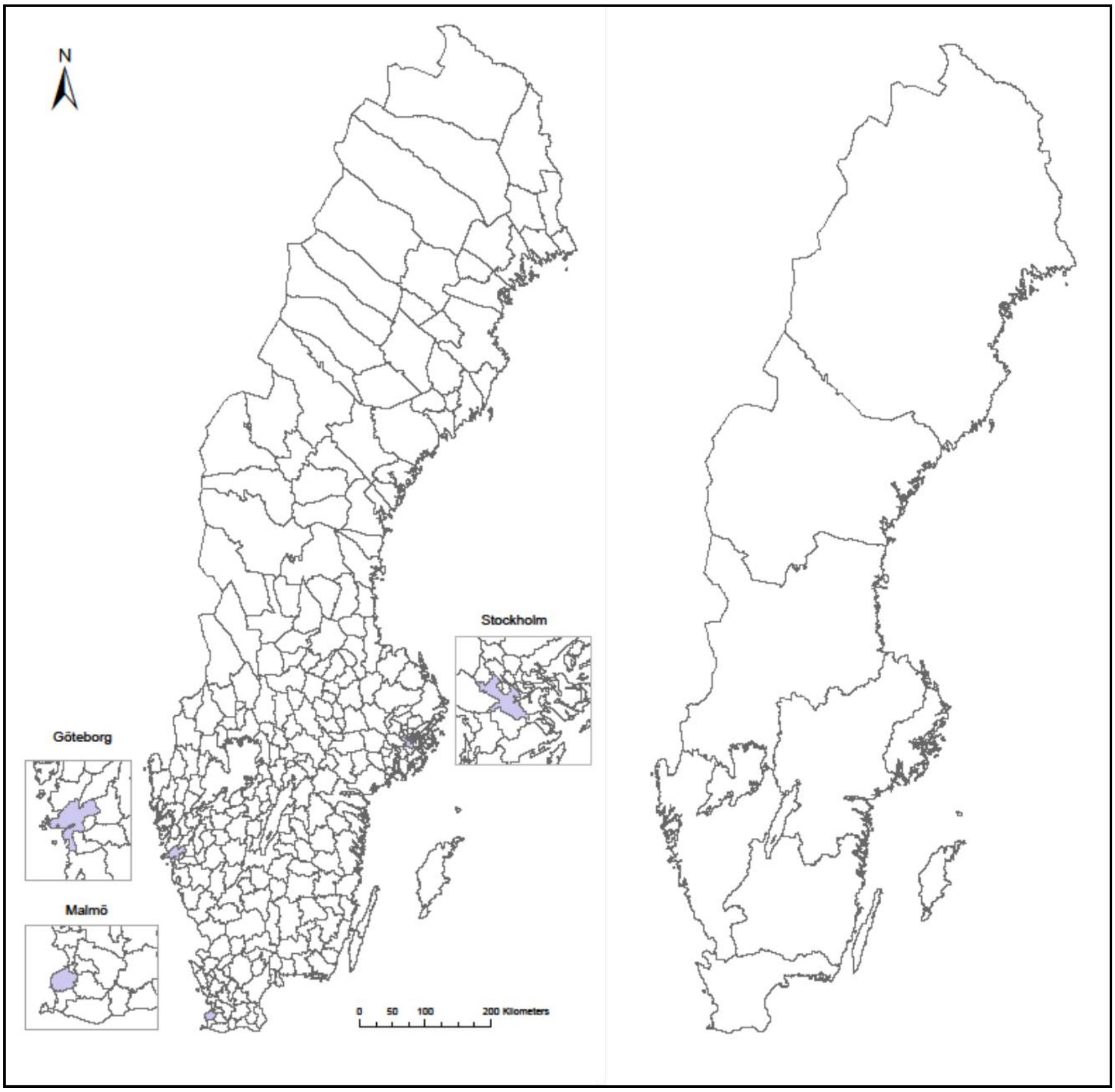

Figure 3 Municipalities (left) and NUTS 2 regions (right) of Sweden. The three metropolitan municipalities and their surroundings inserted

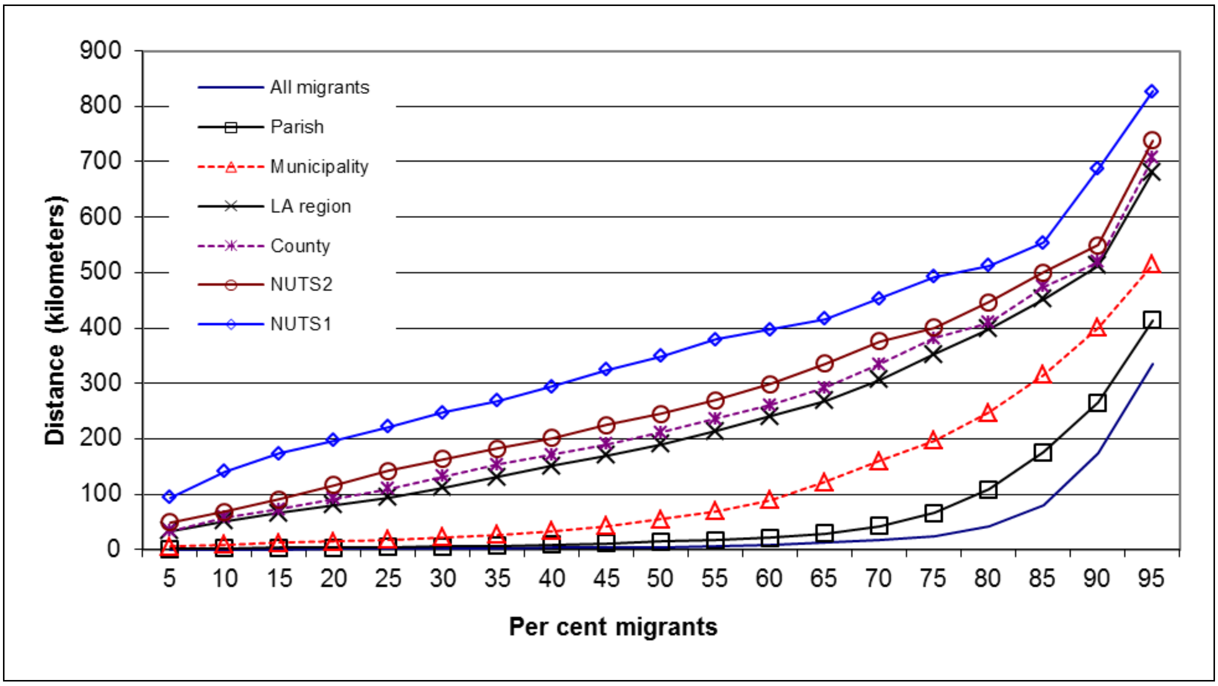

Figure 4 Distribution of migrants across distance and administrative borders, 2008