International Institute for
Applied Systems Analysis
Schlossplatz 1
A-2361 Laxenburg, Austria

Tel: +43 2236 807 342
Fax: +43 2236 71313
E-mail: repository@iiasa.ac.at
Web: www.iiasa.ac.at

Working Paper **WP-17-014**

Bilateral Migration Measures

Wei Qi (qiwei@igsnrr.ac.cn)
Raya Muttarak (muttarak@iiasa.ac.at)
Guy Abel (abel@iiasa.ac.at)

Approved by

Wolfgang Lutz
Program Director, World Population Program
June 2017

Contents

Abstract

The size of migration corridors is typically compared based on the size or rate of the flow. Migration flow counts fail to control for heterogeneity in the population size of regions. Migration rates (or intensities) adjust for population size of either the sending or the receiving region, but not both. To this end, this paper proposes two alternative measures of bilateral migration that express the size of origin-destination corridors in relation to the total migration flows during a period or the total population. The proposed Standardized Migration Affinity (*SMA*) index and Standardized Migration Velocity (*SMV*) index enable us to describe migration systems of different geographical scales and identify prominent migration corridors that are disguised when studying bilateral migrant flow counts or rates. Using examples of internal migration in China, we illustrate the application of the proposed measures at various geographic levels.

Acknowledgments

We acknowledge the support from the Young Scientist Summer Program (YSSP) in providing opportunities for the authors to work together at the International Institute for Applied Systems Analysis (IIASA). We are also grateful to Warren Sanderson for his constructive comments on our work during summer 2016.

About the Authors

Wei Qi is an Assistant Professor at the Institute of Geographic Science and Natural Resource Research, Chinese Academy of Sciences, China

Raya Muttarak is a Senior Lecturer at the School of International Development, University of East Anglia, UK. She is also a Research Scholar at the Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW, WU), Vienna Institute of Demography of the Austrian Academy of Sciences and International Institute for Applied Systems Analysis, Laxenburg, Austria.

Guy Abel is a Professor in the School of Sociology and Political Science and a Leader of the International Migration pillar in the Asian Demographic Research Institute at Shanghai University, China. He is also Research Scholar with the Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW, WU), International Institute for Applied Systems Analysis, Laxenburg, Austria.

Bilateral Migration Measures

Wei Qi
Raya Muttarak
Guy Abel

1 Introduction

Migration is often high on the policy agenda of national governments and non-governmental organisations. A lack of migration data, or poor presentation of such data, often lead to misperceptions about the scale of migration and its effects making the task of decision-makers to develop effective policies more difficult. Poor-quality and scattered information feeds prejudice, stereotyping, and can distort public debate. Without reliable, accessible, and balanced information, the sound management of migration becomes challenging (Laczko 2016).

Typically counts of bilateral migration flows between a set of origin and destination regions are used as a basis for analysis. Comparisons are carried out between regions, summarising the largest sending, receiving regions or migration corridors. However, in most migration systems, the regional population distribution is often uneven and thus simple analyses based solely on summaries of the migration flow counts can be distorted by underling heterogeneity of the populations at risks. One typical approach to address differences in the population sizes of regions is to divide the bilateral migration flow counts by population size of the origin or destination. This approach is useful when only summary measures by region, such as the crude inflow or outflow rate are required. However, if an analyst wishes to compare the strength of migration linkages between regions, rate measures are still likely to mislead as they are standardised to either the population size of the origin or destination region, but not both simultaneously. Current methods for describing and analysing bilateral migration linkages are typically based on interaction terms in fitted regression models. These terms can be difficult to interpret as they conditional on other parameter values, such as a baseline origin-destination pair, to ensure the identifiability.

In order to better describe a system of bilateral migration flows, a measure that accounts for functional linkages between regions and allows for a direct comparison across geographical units and time is required. To this end, this paper proposes two an alternative method to measure bilateral migration. These new methods are an improvement over existing measures of bilateral migration as it 1) captures interactions between places of origin and destination; 2) accounts for origin and destination population size or overall migration levels; and 3) provides a relatively easy to interpret index in relation to other bilateral pairs in the migration systems. Using examples of internal migration in China, we compare and illustrate how migration patterns vary with different indicators of migration flows.

The remainder of the paper is organised as follows. The next section provides a review of existing measures of migration focusing on the indicators that capture links between origin and destination and discusses strengths and limitations of each method. Subsequently, we present an alternative measure of bilateral migration and then demonstrate an application of the measure using interregional migration in China. The final section concludes.

2 Review of Existing Methods

Several measures have been put forward to provide meaningful, comparable indicators of migration. Bell et al. (2002) provided a useful distinction of different measures of migration according to the dimension of measure considered. These dimensions include: 1) measures of the intensity of migration; 2) measures of the distance of migration; 3) measures of migration connectivity; and 4) measures of the effect of migration. The measures of migration intensity are designed to describe the overall level or incidence of mobility capturing both transition probabilities and movement rates. Some of these measures such as the standardized migration probability (*SMP*) or the migration expectancy apply demographic tools such as life-table or age-sex specific migration rate to capture the propensity to migrate. The measures of migration distance summarises the effects of distance across the entire migration system and consider how these vary across space. The measure of migration connectivity captures the strength of the functional linkages between regions i.e. measuring the extent to which regions are connected by migration flows. The measures of migration impact consider the extent to which migration transforms and influences population redistribution. Relative strengths, limitations and utility of each measure are described in detail in Bell et al. (2002).

A series of measures of migration connectivity presented in Bell et al. (2002) primarily measure the degree of connection between different pairs of origins and destinations using the flows between them. These measures essentially describe the roles and functions of individual regions within the multiregional system of migration exchanges and can even be applied to population projections (Rogers & Raymer 1998). However, the measures proposed such as the index of migration connectivity, the index of migration inequality and the Gini index represent only one summary value of all bilateral flows but not the degree of connection between each pair of origin and destination.

Network analysis literature provides a number of methods to measure the centrality of regions (Nogle 1994; Davis et al. 2013). Different centrality measures provide different summaries on each region based on weights, for which the size of the bilateral migration flows can be used. For example, degree centrality calculates high values for regions with large sum of inflows and outflows. Closeness centrality quantifies how central or peripheral a region is by calculating higher values to those with large flow between other high ranking regions. Such statistics provide summaries for each region, by aggregating over the bilateral information. For large systems, they can be valuable summaries, but centrality measures are also unable to distinguish between regions with different population sizes.

In this paper we focus specifically on the measures of bilateral migration that can control for the population size in both the origin and destination regions. There are a number of measures that explicitly summarise the two-way migration interaction between origin and destination, besides the migration flow count.

As previously discussed, one approach to address differences in the population sizes of regions is to calculate the **migration rate** by dividing the bilateral migration flow counts (M_{ij}) by population size in the origin (P_i)

$$m_{ij} = \frac{M_{ij}}{P_i}$$

or destination (P_j):

$$m_{ij} = \frac{M_{ij}}{P_j}$$

These measures only standardise the migration flow to either the population size of the origin or destination region, but not both simultaneously. In order to address this issue, one available option is to divide the migration flow count by the product of the origin and destination populations.

$$m_{ij} = \frac{M_{ij}}{P_i P_j}$$

Fielding (1992) calls this measure a **'migration velocity'**. While the population sizes at both origin and destination are standardised in this measure, it does not capture the magnitude of the links between regions relative to other bilateral pairs. Migration velocity in a migration corridor can be high because of strong push or pull effects that influence all of the bilateral flows to or from that particular region.

In regional science, an **interaction value** (IV) has long been used to define travel to work areas (Smart 1974; Ball 1980);

$$m_{ij} = \frac{M_{ij}^2}{O_i I_j}$$

where $O_i = \sum_j M_{ij}$ are the outflows from each region and $E_j = \sum_i M_{ij}$ are the inflows to each region. This measure has also been applied to describe and predict migration origin-destination flow tables (Rogers et al. 2002) Unlike the migration velocity measure, controls for the push and pull effects are used to standardise the flow count.

Rogers et al. (2002) demonstrate the use of a **modelling decomposition approach** to analyse migration patterns. Theirs, as with other similar analyses, form a measure for interaction between regions by controlling for push and pull effects. These can be written in various guises, but typically involve the regression of the migration flow counts (or a transformation of the counts) on categorical explanatory variables for each origin and destination region and an interaction term.

$$\log M_{ij} = \beta^C + \beta_i^O ORIG_i + \beta_j^D DEST_j + \beta_{ij}^{OD} ORIG_i{:}DEST_j + \varepsilon_{ij}$$

This results in a four sets of parameters to provide a complete description of a migration system. The β^C represents the overall level of migration. The β_i^O and β_j^D terms represent the push and pull effects for each region, with the exception of a reference region, whose parameter value might be one or zero or some other values, depending on a form of the dependent variable and a parameter coding scheme being used. The set of values β_{ij}^{OD} represent the strength of connection between regions in relation to a reference migration corridor. The β_{ij}^{OD} terms provide a strict indication of the linkages between regions having controlled for the overall level of flows and the size of migration in and out of each region. Rogers et al. (2002) used a regression model approach, assuming the dependent variable was the natural logarithm of migration flow counts and the errors with a Poisson distribution. As a result, the Poisson log-linear model provides a saturated fit to each observation (i.e. there are no remaining degrees of freedom). The parameters in the model can be calculated directly when using the authors total reference coding system, where, for example, the interaction terms can be obtained as:

$$\beta_{ij}^{OD} = \frac{M_{ij}}{\beta^{C} \beta_{i}^{O} \beta_{j}^{D}}$$

The resulting parameter estimates are constrained to provide an intuitive interpretation of their values without reference to a single migration corridor. When β_{ij}^{OD} is much greater than one, it signifies that there is a strong association between regions, even after controlling for the total inflows and outflows, whereas ratios less than one indicate the opposite (i.e. a weaker migration linkage than expected in an independence model). Raymer et al. (2006) extended this approach to include an age dimension to the existing origin and destination dimension, for analysis and projection of migration rates in age-specific bilateral migration flows.

Other modelling approaches have been proposed that are related to Zipf's (1946) gravity model, where flows are dependent on the population size at the origin and destination and the distance between each region:

$$\log M_{ij} = \beta_1^O \log P_i + \beta_2^D \log P_j - \beta_3 \ \log DIST_{ij} + \varepsilon_{ij}$$

The error terms in this model can be used to study difference between observed migration flows and those expected from a gravity model (Shen 2016). The parameter terms can be estimated using standard linear regression procedures or they can be fixed to $\beta_1^O = \beta_2^D = 1$ and $\beta_3 = -1$ (and ε_{ij} set to zero) as in Zipf's original specification.

Both of these model-based approaches provide measures of bilateral associations between regions. Whilst the model decomposition approach using the total reference coding scheme provides intuitive interaction terms, it does not control for population size. The error-based approach can control for population in both the origin and destination regions but provides less interpretable measures of migration linkages between regions.

3 Alternative Methods to Measure Migration

In this paper, we propose two alternative methods to measure bilateral migration connectivity which considers: a) the strength of association between origin and destination areas; and b) population size at both the origin and destination.

3.1 Standardized Migration Affinity Index (*SMA* Index)

Migration flow counts are affected by characteristics of a geographic unit. Geographical areas with a large population size are not directly comparable to units with a small population size, a form of the modifiable areal unit problem (*MAUP*) (Openshaw & Taylor 1979). Thus, a migration flow count itself can not reveal the intrinsic connectivity between the origin and destination. This is problem faced when making comparison of other measures across different geographical areas. For example, in the assessment of the level of economic development the Gross Domestic Product (GDP) per capita is more suitable than total GDP.

In order to provide an easy to interpret measure of the migration relationships that can account for the push and pull effects of a given region, we propose a new **'Standardised Migration Affinity'** (*SMA*) index[1]:

$$SMA_{ij} = \frac{M_{ij}^2}{O_i I_j} \frac{1}{\overline{IV}}$$

As with the interaction value (*IV*), the *SMA* value of each migration corridor controls for regions with large inflows and outflows, so that those that are more concentrated have higher values. Migration corridors with high levels of affinity have greater levels of migration than expected under an independence model, where all moves between regions are random, having controlled for the overall push and pull factors to and from each region. Our *SMA* index extends the *IV* values by standardising all values to the average *IV*:

$$\overline{IV} = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{M_{ij}^2}{O_i I_j}$$

where *n* denotes the count of a spatial unit and *N* denotes the number of migration flows. If we only consider inter-units migration flows, then $M_{ii}=M_{jj}=0$ and $N=n^2-n$. If inner-unit migration flows are also considered, then M_{ii} and M_{jj} are non-zero and $N=n^2$.

Our standardisation allows for a clear interpretation and facilitates comparisons between migration systems. When the SMA_{ij} is greater than one, a higher migration affinity than the average level of the network is present. When the level of SMA_{ij} is less than one, a relatively low migration affinity is present. Furthermore, the sum of SMA_{ij} from origin *i* or to destination *j* denotes the total in-migration affinity as an origin or total out-migration affinity as a destination.

3.2 Standardised Migration Velocity Index (*SMV* Index)

While *SMA* index considers the influence of both total out-migrants at the origin and total in-migrants at the destination, it does not account for the population sizes in each region. The ratios of migrants to the total population or to the non-migrants indicate migration

[1] Note that 'affinity' is frequently used in sociology to refer to inter commonalities in close groups whereby people in the same social circle (e.g. religion, occupation, sport club) share similar qualities, ideas and interests (Moreland & Beach 1992). However, in this context, 'migration affinity' stands for the connectivity between two places.

intensity. When two migration corridors have the same numbers of migrant flows and the same *SMA* values, but different numbers of non-migrants in the origin (or the destination), there exists different probabilities of individuals to migrate through each migration corridor. Accordingly, we propose a **'Standardized Migration Velocity'** (*SMV*) index to measure the migration interaction:

$$SMV_{ij} = \frac{M_{ij}^2}{P_i P_j} \frac{1}{\overline{MMV}}$$

As with migration velocity values, the *SMV* index of each migration corridor controls for migration between regions with large population sizes, so that large flows between smaller regions are given a higher weight. Unlike the migration velocity measure, we use the square of the migration count as the numerator. This adds a higher weight to larger flows. We also standardise in order to provide a more intuitive value using our modified migration velocity (*MMV*):

$$\overline{MMV} = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{M_{ij}^2}{P_i P_j}$$

Similar to the *SMA* index, SMV_{ij} above or below one indicate a higher or lower migration velocity comparing to the average value. The sum of SMV_{ij} across *i* or *j* denote the total in-migration velocity in the origin or total out-migration velocity in the destination, respectively.

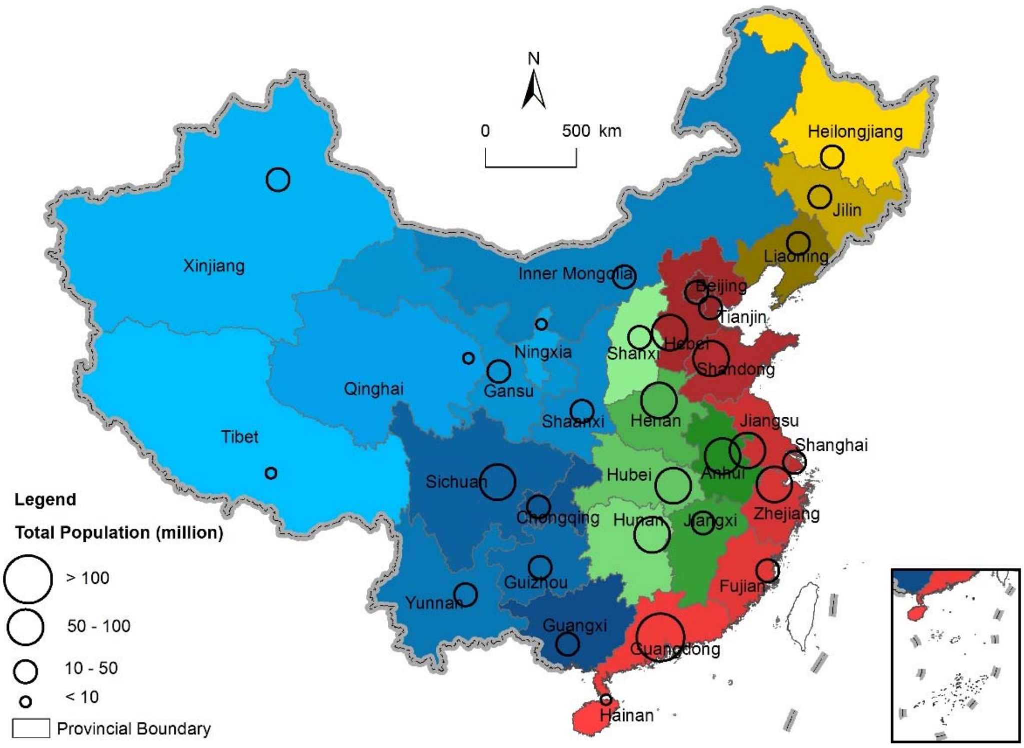
4 Calculating Different Measures of Bilateral Migration Using Example of Interprovincial Migration in China

In this section we use internal migration in China to illustrate the *SMA* and the *SMV* indices. The Population Census Office (2012) provides a population origin-destination migration flow table during 2005-2010 among 31 provincial units in mainland China (Hong Kong and Macau are not included) from the 2010 Census. Destination in the table is defined as the current place of residence while origin is defined as the residential place five years ago. The question on the previous place of residence in the Census, used to enumerate migration flows, was asked only to the population aged five years and above. People who were born or died during the past five years were not included.

As shown in Figure 1, the 31 provinces have different population sizes and land sizes. For example, Tibet has a large land area but has less than 10 million inhabitants as compared to Beijing which has much smaller land area but with the population size of over 25 million.

Following the regional development policy in China, the 31 provincial units can be divided into four regions: the East, the Centre, the West and the Northeast (Liu & Xu 2017). To demonstrate and simplify the calculation of *SMA* value and *SMV* value, we aggregated inter-provincial migration table into inter-regional migration origin-destination flow table. Thus, the inter-provincial migrants within the same region, which is called inner-regional migrants, are not regarded as inter-regional migrants in this exercise.

Figure 1: Spatial distribution of 31 provincial units and their populations in mainland China in 2010



Notes: The colour represents each provincial unit. The same colour series indicate that the provincial units belong to the same region. The red series, green series, blue series and yellow series stand for the East, the Centre, the West and the Northeast, respectively. The provincial total population data are obtained from residential population census data in 2010.

The migration flows from the Centre to the East and from the West to the East were the largest during the period. Flows in and out of the Northeast were relatively lower because of its smaller geographic and population size. The East gained the highest number of migrants whilst the Centre lost the most migrants. Only the Eastern region grew in population size as a result of migration, from 429.59 million to 458.74 million, whilst the three other regions all experienced depopulation.

The *SMA* index is shown in the middle panel of Table 1 along with the origin and destination sums. There are only four migration corridors that have relatively higher migration affinity (*SMA* value above 1). In most other migration corridors the migration affinity is well below unity. The largest migration corridors are those to the East (from the Centre and West) and from the East (to the Centre and West). The Centre has the largest origin sum of SMA values, whilst the East has the largest destination sum of *SMA* values.

Table 1: Various measures of migration flows in China during 2005-2010.

Migration Flows (million)

Origin	Destination					Population 2005	Population 2010
	East	Centre	West	Northeast	Sum		
East		1.84	1.79	0.44	4.08	427.59	458.74
Centre	20.45		1.75	0.31	22.51	332.83	313.18
West	12.99	0.88		0.33	14.20	338.52	328.18
Northeast	1.80	0.13	0.31		2.25	99.99	98.82
Sum	35.24	2.86	3.86	1.08	43.03	1198.93	1198.93

SMA Index

Origin	Destination				
	East	Centre	West	Northeast	Sum
East		2.29	1.61	0.35	4.25
Centre	4.15		0.28	0.03	4.46
West	2.65	0.15		0.66	2.86
Northeast	0.32	0.02	0.09		0.43
Sum	7.12	2.47	1.97	0.44	12.00

SMV Index

Origin	Destination				
	East	Centre	West	Northeast	Sum
East		0.08	0.07	0.01	0.16
Centre	8.23		0.08	0.01	8.31
West	3.26	0.02		0.01	3.30
Northeast	0.21	0.002	0.01		0.22
Sum	11.71	0.10	0.16	0.03	12.00

The inter-regional population migration flow table is given in the top panel of Table 1. This matrix only considers inter-regional origin-destination migrants' volume. Each row indicates the origin region and each column represents the destination region. Natural population growth is not considered in this table, thus the total population in 2005 and in 2010 is identical

Here, we choose two migration corridors; 1) the East-West and 2) the Centre-West to demonstrate the function of *SMA* index. Although the original migrants counts of these two flows are almost identical (1.79 million and 1.75 million), their *SMA* index values are quite different. The *SMA* value of the former (1.61) is significantly higher than that of the latter (0.28). Despite of the closer geographical proximity of the Centre to the West, the *SMA* index illustrates the much stronger connection of the West to the East, due to its higher level of economic development. The higher affinity in the East-West corridor (a higher migration flow than under expected in an independence model) is hidden when considering the raw flow count.

The *SMV* index is shown in the bottom panel of Table 1. Since migration flows are measured during the period 2005 to 2010, when calculating *SMV* values, we used the population in 2005 as the origin's population and took the population in 2010 as the destination's population. The sum value of *SMV* in each region is also given, showing that the Centre and the East are the main origin and destination, respectively comparing to other regions. There are only two migration corridors that have a higher than average migration velocity; the Centre-East and West-East. All other migration corridors have *SMV* values less than 0.05, illustrating the dominate role of relative roles (adjusting for the population size), of the Centre and West to East migrant corridors in the Chinese migration system during 2005-10.

The migration affinity, measured using the *SMV* index, can differ greatly from the *SMA* index. For example, based on the *SMA* values migration corridors to the North-East varied greatly (0.35 from the East, 0.03 from the Centre and 0.66 from the West) whereas the *SMV* values were all very similar (close to 0.01). These differences illustrate the divergent uses of the indices in addressing different questions. If the relative size of the population and the probability of movements are of interest, the *SMV* index provides a useful comparative measure. If the relative size of the migration flow with respect to all migrant corridors is of interest, the *SMA* is of more relevance.

5 Extension of SMA and SMV to Include Within Region Migration

When different geographic scales are combined, both inter-regional migration and inter-provincial migration are usually considered together. For example, when we transfer inter-provincial migration into inter-regional migration, the inter-provincial migration can be classified into two categories: inter-regional migration and intra-regional migration between different provincial units. The above example (in Table 1) only considers inter-regional migration while intra-regional migration between different provincial units are classified as no migration. However, if the study is also interested in intra-regional migration, the inter-provincial population migration within the same region should be considered as a migration flow.

Table 2 presents migration flows in China including intra-regional migration. While the off-diagonal values in the top panel of Table 2 are the same as in the top panel of Table 1, the diagonal values now represent the within region inter-province migration flow counts. The calculations of the *SMA* and *SMV* values can be easily altered to account for migration flow counts in the diagonal cells. The total number of bilateral migration corridors becomes n^2, not n^2-n. The results for *SMA* and *SMV* indices are shown in the middle and bottom panel of Table 2, respectively.

Table 2: Various measures of migration flows in China during 2005-2010 including within region flows

Migration Flows (million)

Origin	Destination					Population 2005	Population 2010
	East	Centre	West	Northeast	Sum		
East	7.17	1.84	1.79	0.44	11.25	458.74	427.59
Centre	20.45	1.12	1.75	0.31	23.63	313.18	332.83
West	12.99	0.88	2.92	0.33	17.11	328.18	338.52
Northeast	1.80	0.13	0.31	0.75	3.00	98.82	99.99
Sum	42.41	3.98	6.77	1.83	54.99	1198.93	1198.93

SMA Index

Origin	Destination				
	East	Centre	West	Northeast	Sum
East	1.51	1.06	0.59	0.13	3.29
Centre	5.84	0.19	0.27	0.03	6.33
West	3.25	0.16	1.03	0.05	4.49
Northeast	0.36	0.02	0.07	1.44	1.89
Sum	10.96	1.43	1.95	1.65	16.00

SMV Index

Origin	Destination				
	East	Centre	West	Northeast	Sum
East	0.95	0.09	0.08	0.02	1.15
Centre	9.95	0.04	0.10	0.01	10.11
West	3.95	0.03	0.28	0.01	4.27
Northeast	0.26	0.002	0.01	0.21	0.48
Sum	15.11	0.16	0.47	0.25	16.00

The SMA and SMV indices accounting for intra-regional migration are presented in the middle and bottom panels in Table 2, respectively. The intra-regional flow in the Centre (1.12 million), for example, was much larger than that in the Northeast (0.75 million). However, when *SMA* or *SMV* measures are applied, the intra-regional moves in the Northeast appeared to have a stronger migration affinity or velocity than that in the Centre. Here we are able to present how active the intra-provincial migration in the Northeast is, which is not possible to capture using migration flow count alone. In general, during 2010-2015, six flows (three inter-regional flows and three intra-regional flows) had a *SMA* value above one. Based on the SMV index, only the inter-regional flows from the Centre and the West to the East have SMV value above one. The SMV index based on the inter-regional moves illustrates the relatively minor scale of the intra-regional inter-provincial flows once the population at risk was considered.

There are also major differences between the patterns of inter-regional migration without intra-regional migration in Table 1 and inter-regional migration with intra-regional migration in the diagonal cells in Table 2. The Centre-East and West-East are the two largest migrant flow corridors based on raw-flow counts. However, the intra-regional flows in the East (7.17 million) and in the West (2.92 million) became the third and the fourth largest migration corridors. The *SMA* and the *SMV* patterns also altered. For example, the original inter-regional pattern showed that the out-migration *SMA* sum value in the East (4.25) was higher than that in the West (2.86), but in Table 2 where intra-regional migration is included, the West (4.49) has a higher out-migration *SMA* sum value than the East (3.29). Similarly, the in-migration *SMV* value (0.10) in the Centre was larger than that in the Northeast (0.03) when interregional flows are ignored (Table 1), but smaller when included (0.25 in the Northeast and 0.16 in the Centre; Table 2). This example shows that the flow pattern varies not only by different measures but also by different definitions of migration.

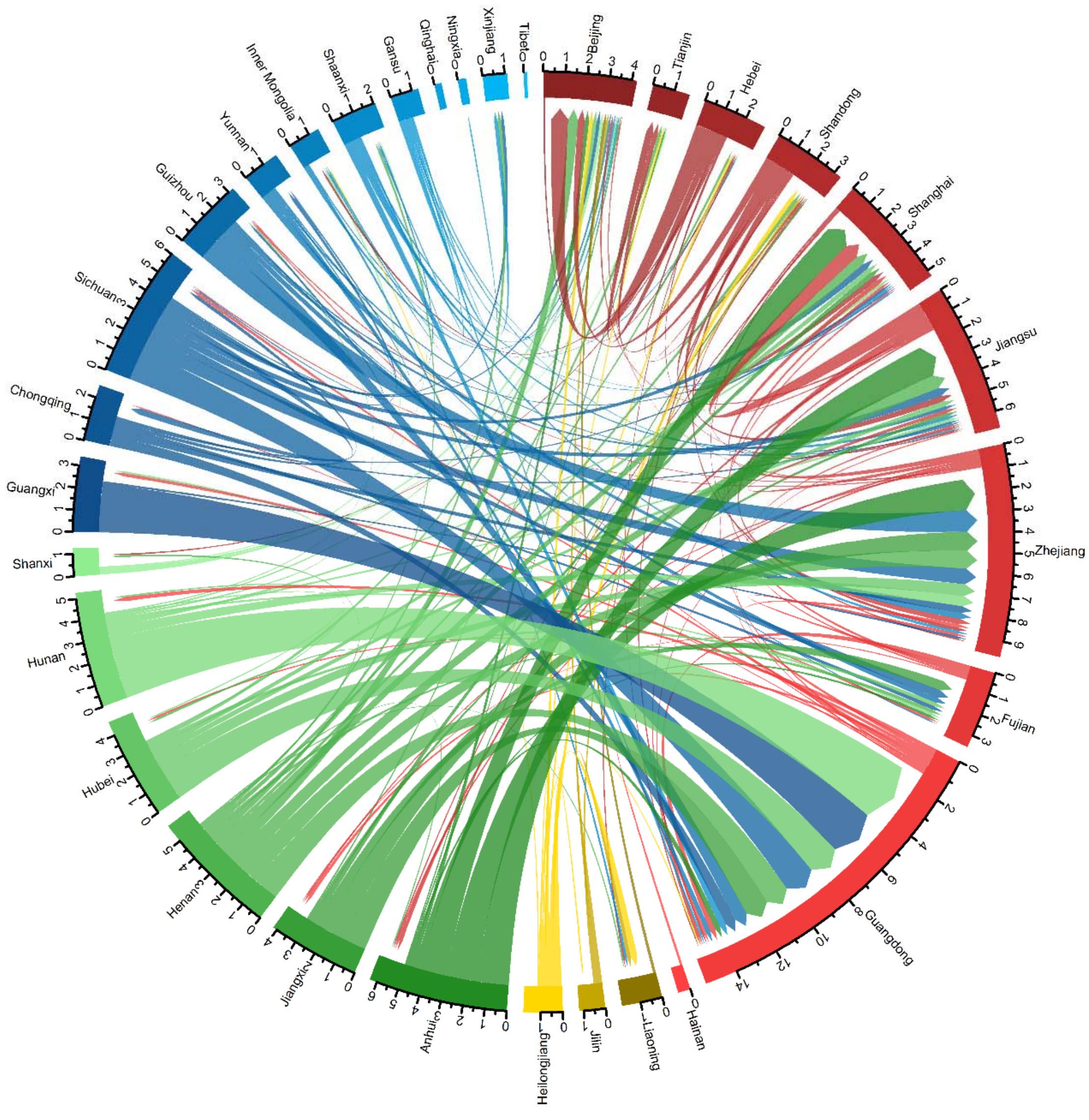
6 Extension of SMA and SMV to Describe China's Inter-Provincial Population Migration

To explore the function of the *SMA* and the *SMV* for different geographic scales, we further apply the *SMA* and the *SMV* to describe China's inter-provincial population migration during 2005-2010. Since the opening up and market-oriented reforms in 1978, China's economy developed rapidly, especially in the coastal provinces such as Shanghai, Jiangsu, Zhejiang, Guangdong and Fujian (Bao et al. 2002). There is a significant difference in geographic feature and regional economic performance between the coastal provinces and inner land provincial units (Wei 1999). The growth of manufacturing employment in the coastal regions on the one hand and the rural surplus labour coupled with the widening urban-rural standard-of-living gap on the other paved the way for the constant rise in internal migration in mainland China (Liang 2001). Millions of people from the rural areas and interior regions move to the coastal regions in search of better life chances. Geographical mobility in China is indeed one of the most active and the largest internal migration systems in the world (Liu et al. 2011).

Here, we use a chord diagram plot to visualize the interprovincial *SMA* index and *SMV* index results for migration in mainland China. As shown in Figures 2, 3 and 4, different colours represent different provinces with common colour palettes for provinces in the same region. The arrow shows the flow direction from the origin to the destination. The width of each flow indicates the size of the bilateral migration measure and the axes show the sum values into and out of each province (Qi et al. 2017). The flow system of inter-provincial population migration, which includes 930 flows, is much more complex than that of inter-regional population migration presented in Section 5. Therefore, in Figure 2 only the flows above the average level and in Figures 3 and 4 the migration corridors with SMA and SMV values above one are drawn in the circular plot, whilst those less than one are not shown.

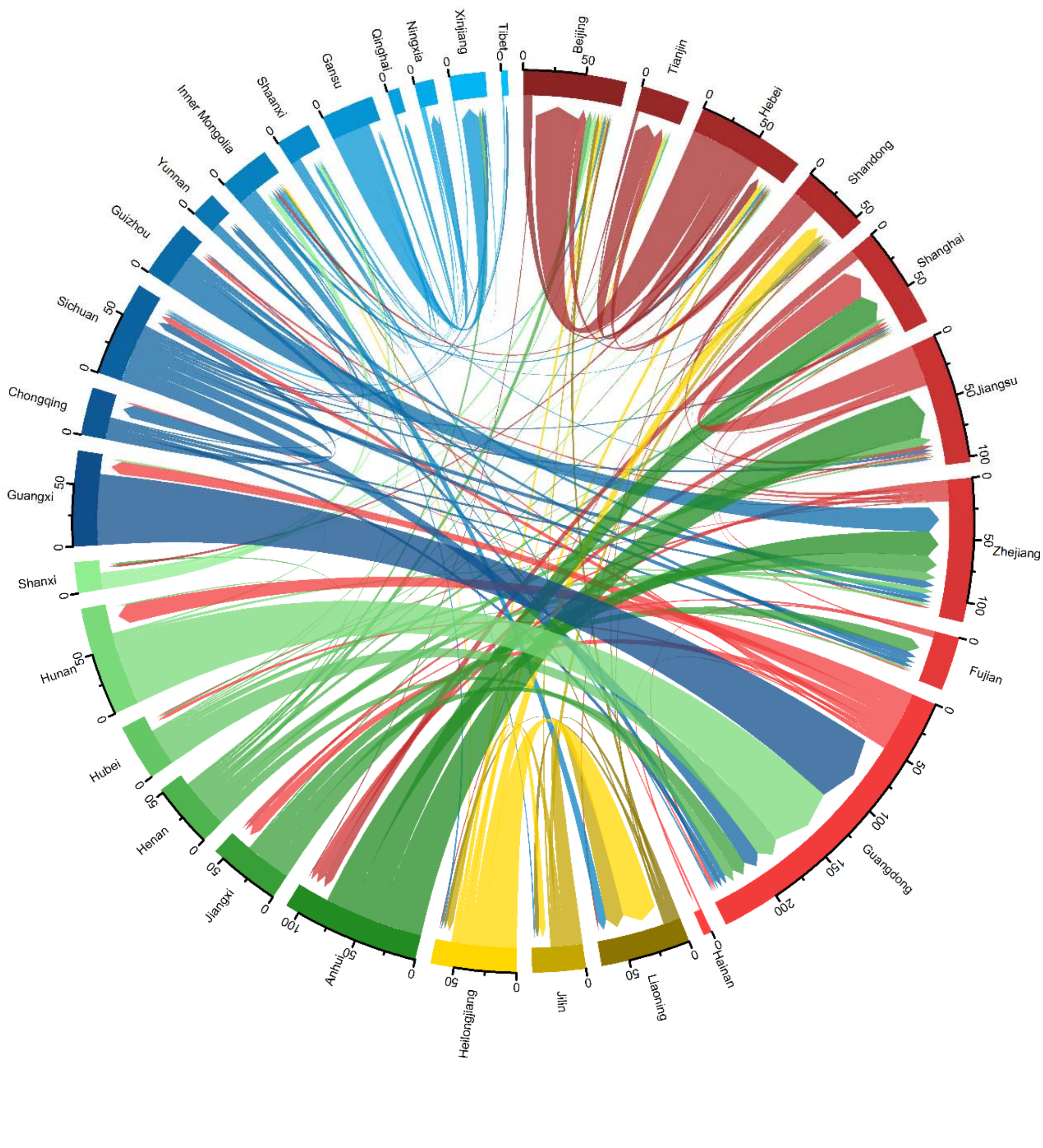
The migration flow values presented in Figure 2 are comprised of 158 corridors above the average level (0.057 million), which are presented in the plot. The relatively larger flows are mostly from the provinces located in the West or the Centre to those provinces in the East. The largest one is from Hunan to Guangdong, with 2.93 million migrants during 2010—2015. The provinces located in the East such as Guangdong, Zhejiang and Jiangsu are the main migration destinations. On the other hand, the provinces in the Centre or the West, such as Anhui, Jiangxi, Henan, Hubei, Hunan and Sichuan, are the major migrants' sending-out origins.

Figure 2: Circular plot of raw values of China's inter-provincial migration



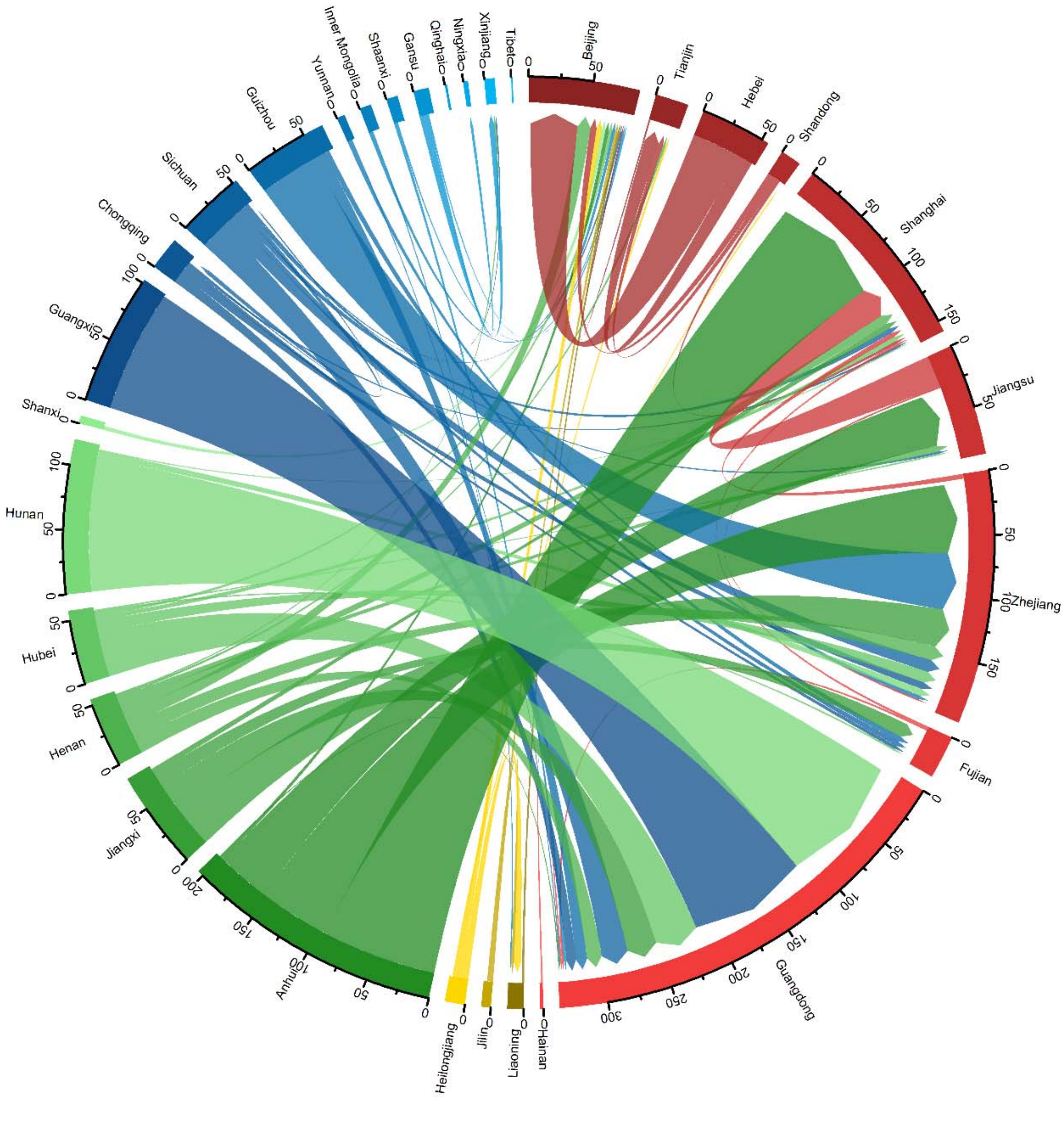
The *SMA* plot shown in Figure 3 is based on 137 corridors with values above one. Whilst the largest flow count is from Hunan to Guangdong and the second largest from Guangxi to Guangdong, the ranking is inverted when based on the *SMA* index. For the in-migration to Guangdong, Guangxi has stronger affinity than Hunan. The small flows such as from Gansu to Xinjiang, from Heilongjiang to Liaoning and from Hebei to Beijing become stronger based on the *SMA* index. Small out-migration flows from the provinces in the East to the Centre or the West such as those from Guangdong became more prominent in the *SMA* plot.

Figure 3: Circular plot of S*MA* values of China's inter-provincial migration



The *SMV* plot presented in Figure 4 is based on 71 flows with the values above one. Similar to migration flow count and the SMA index, migration from Hunan to Guangdong remains the largest corridor based on the SMV index. But the *SMV* value appears to be much stronger than its count value. Comparing to the flow count and the *SMA* value, the *SMV* value of the Anhui to Shanghai corridor becomes greater than that from Anhui to Jiangsu. Furthermore, the SMV index allows us to identify a truly popular migration destination such as Shanghai. The sum value of in-migration *SMV* in Shanghai is larger than that in Jiangsu. This pattern cannot be captured using migration flow count or SMA index since Shanghai has a relatively small population size – only 30% that of Jiangsu. Through adjusting the flow value by population size, the *SMV* index highlights the importance of Shanghai as an active in-migration destination.

Figure 4: Circular plot of *SMV* values of China's inter-provincial migration.



The sizable migration flows between populous provinces are clearly evident when studying the inter-provincial migration in China especially from the provinces in the Centre to the provinces in the East. The *SMA* index and the *SMV* index provide an alternative view of migration patterns that cannot be captured using migration flow counts alone. For example, the *SMA* index highlights migration affinity from Heilong to Liaoning within Northeast China, which is usually negligible when compared to the large inter-provincial migration flows to the developed Eastern area. Likewise the *SMV* index, which adjusts for the population size of both migrants and non-migrants, enables us to identify Anhui as a migrant-sending province and Shanghai as a migrant-receiving province – patterns that cannot be derived directly from either the migration flow count or the *SMA* index.

7 Discussion and Conclusion

The standardized migration affinity (*SMA*) index and the standardized migration velocity (*SMV*) index introduced in this paper provide alternative measures to describe the bilateral migration patterns of a migration system. In our proposed SMA index we use the total out-migrants at origin and total in-migrants at destination to capture affinity between the two places. In our *SMV* index we adjust for total population sizes at both the origin and destination to address differential rates of migration from heterogeneous population sizes. The *SMA* index highlights the largest and smallest corridors in a migration system while the *SMV* presents the activity of the whole population in relation to each migration corridor.

Both new measures improve our understanding of the complex population migration system by going beyond simple counts of bilateral migration flows. Using standardisations, our proposed indices can be interpreted using the average value of one as a benchmark. This allows straightforward comparisons of migration corridors to the system as a whole and the relative importance of origin and destination regions by comparing the summed *SMA* or *SMV* to the number of regions. Both the *SMA* index and the *SMV* index can be applied to different geographic scales such as inter-regional and inter-provincial migration and to different definitions of migration such as inter-regional migration with or without considering intra-regional migration. They could also be used to provide potential comparisons between internal migration corridors in different countries.

This paper proposes two alternative measures of bilateral migration flows which extend beyond a traditional measure of bilateral migration using migration flow count. Migration flow counts are affected by the statistical spatial unit scale and population size at the origin and destination. The migration flow from one region to another may appear large simply because the origin has large population size and consequently greater number of the population at risk of migration. Alternatively, migration flow from one area to another may appear small because the sending area has small population size. Once the overall level of migration is adjusted for i.e. accounting for the proportion of inflows and outflows for each origin-destination pair or the population size it is possible to detect migration connectivity between a specific origin and destination. This information can aid policy makers in understanding and predicting important bilateral migration flows which are not obvious using only the migration flow counts.

We have shown that alternative measures of bilateral migration such as our *SMA* index and the *SMV* index server to highlight important features of a migration system. In the case of China, the ongoing massive influx of rural migrants into the urban areas raised concerns for the Chinese government about over-urbanisation and the strain of the population growth on the urban infrastructure and facilities (Chai & Karin Chai 1997). In March 2014, the Chinese government launched the *New Urbanisation Plan* aiming to pursue a stable and balanced version of urbanisation and regulate population migration, among other things. Both *SMA* and *SMV* indices can potentially highlight origin, destination and migrant corridors when measuring the impact of new policies.

8 References

Ball, R.M. 1980. The use and definition of Travel-to-Work Areas in Great Britain: Some problems. *Regional Studies* 14(2): 125–139.

Bao, S. et al. 2002. Geographic factors and China's regional development under market reforms, 1978–1998. *China Economic Review* 13(1): 89–111.

Bell, M. et al. 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.

Chai, J.C.H. and B. Karin Chai. 1997. China's floating population and its implications. *International Journal of Social Economics* 24(7/8/9): 1038–1051.

Davis, K.F. et al. 2013. Global Spatio-Temporal Patterns in Human Migration: A Complex Network Perspective. *PLoS ONE* 8(1).

Fielding, A.J. 1992. Migration and social mobility: South East England as an escalator region. *Regional Studies* 26(1): 1–15.

Laczko, F. 2016. Improving Data on International Migration and Development : Towards a Global Action Plan ? "Improving Data on International Migration - towards Agenda 2030 and the Global Compact on Migration ." Pages 1–12 in

Liang, Z. 2001. The Age of Migration in China. *Population and Development Review* 27(3): 499–524.

Liu, S. et al. 2011. The regional types of China's floating population: Identification methods and spatial patterns. *Journal of Geographical Sciences* 21(1): 35–48.

Liu, Y. and W. Xu. 2017. Destination Choices of Permanent and Temporary Migrants in China, 1985–2005. *Population, Space and Place* 23(1): n/a-n/a.

Moreland, R.L. and S.R. Beach. 1992. Exposure effects in the classroom: The development of affinity among students. *Journal of Experimental Social Psychology* 28(3): 255–276.

Nogle, J.M. 1994. The Systems Approach to International Migration: An Application of Network Analysis Methods. *International Migration* 32(2): 329–342.

Openshaw, S. and P.J. Taylor. 1979. A million or so correlation coefficients: three experiments on the modifiable areal unit problem. *Statistical applications in the spatial sciences* 21: 127–144.

Population Census Office. 2012. *Tabulation on the 2010 Population Census of the People's Republic of China*. Beijing: Population Census Office under the State Council and Department of Population and Employment Statistical National Bureau of Statistics.

Qi, W. et al. 2017. Circular visualization of China's internal migration flows 2010–2015. *Environment & Planning A.*

Raymer, J., A. Bonaguidi, and A. Valentini. 2006. Describing and projecting the age and spatial structures of interregional migration in Italy. *Population, Space and Place* 12(5): 371–388.

Rogers, A. et al. 2002. Describing migration spatial structure. *Papers in Regional Science* 81(1): 29–48.

Rogers, A. and J. Raymer. 1998. The spatial focus of US interstate migration flows. *International Journal of Population Geography* 4(1): 63–80.

Shen, J. 2016. Error Analysis of Regional Migration Modeling. *Annals of the American Association of Geographers* 106(6): 1253–1267.

Smart, M.W. 1974. Labour market areas: Uses and definition. *Progress in Planning* 2: 239–353.

Wei, Y.D. 1999. Regional inequality in China. *Progress in Human Geography* 23(1): 49–59.

Zipf, G.K. 1946. The P1 P2/D hypothesis: On the intercity movement of persons. *American Sociological Review* 11(6): 677–686.