Title page

Article type: full research manuscript

Comparing intermigration within the European Union and China: an initial exploration

Jianquan Cheng[1*], Craig Young[1], Xiaonan Zhang[2] and Kofi Owusu[1]

[1]School of Science and the Environment, Division of Geography and Environmental Management, Manchester Metropolitan University, John Dalton Building, Chester Street, Manchester, M1 5GD, UK. Tel. 00-44-(0)161-247-1576; fax: 00-44-(0)161-247-1633; j.cheng@mmu.ac.uk; c.young@mmu.ac.uk

[2]Department of Urban Planning and Design, Xi'an Jiaotong - Liverpool University, 111 Renai Road, Suzhou, China, 215123; Tel. 00-86-(0)512-88161771; xiaonan.zhang@xjtlu.edu.cn

[*]School of Science and the Environment, Manchester Metropolitan University, John Dalton Building, Chester Street, Manchester, M1 5GD, UK. Tel. 00-44-(0)161-247-1576; fax: 00-44-(0)161-247-1633; j.cheng@mmu.ac.uk

Abstract

Labour mobility has been extensively studied in China and the European Union (EU). However, there has been very little attempt to compare inter-state migration within the EU and inter-provincial migration in China. This paper provides an account of an initial exploratory quantitative comparison of EU and Chinese intermigration. The paper first makes the case for comparing the EU and China in the context of the growing literature on international comparisons of migration. Problems of data and definition are then explored, and a review of the literature undertaken to identify which determinants could be used as the basis of comparative study. Mobility patterns are measured by comparing selected indicators including inequality, dispersion and effectiveness. The migration process is then measured in terms of the elasticities of intermigration responses to various contributing factors, among which spatial adjacency, distance, economic prospects, migrant networks, and migration policy are filtered out and employed for the study. Gross migration flow models are calibrated to produce initial comparative results, using fixed-effect negative binomial regression methods and a variety of sources of data. The conclusion discusses how this initial exploration has helped identify some potential research directions for future work.

Keywords: China, comparative analysis, determinants, European Union, intermigration, model, pattern

Funding statement

This work was supported by the UK Economic and Social Research Council (ESRC) and the Royal Academy. This paper is one of the research outcomes of the project

‘Internal mobility and integration in the EU and China’ funded by the CO-REACH programme, promoting research cooperation between Europe and China (August 2009–January 2011), grant number ERA-NET CO-REACH-64-028.

Acknowledgements

We would like to thank the anonymous referees for their useful and constructive comments. Particular thanks also go to Bill Jordan and Dita Vogel for constructive and critical comments on an earlier version of this paper.

Comparing inter-migration within the European Union and China: an initial exploration

1. Introduction

International comparative studies of migration have increased in recent years, especially because 'comparative studies can challenge accepted and conventional wisdoms and lead to innovative new thinking' (Bloemraad 2013: 41). However, this remains an under-developed area and in this paper we extend analysis through an initial comparative exploration of inter-migration in China and the European Union (EU). There is a substantial literature on inter-state migration in the EU (Lewer and Van den Berg 2008; Beine et al. 2007; Hooghe et al. 2008; Marques 2010; Pedersen et al. 2008; Warin and Svaton 2008) and inter-provincial migration in China (Cai and Wang 2003; Fan 2005b; Poncet 2006; Bao et al. 2011; Shen 2012) but there have been no previous attempts to compare migration within them (and only one previous qualitative comparison of migrants' social rights in the EU and China (Kovacheva et al. 2012)). As the EU will be treated as one geographic unit in order to compare it with China, what is usually termed international migration from one EU member state to another is here called inter-state migration and inter-province and inter-state migrations are referred to as inter-migration (Wu and Yao 2003) (though this does not include internal migration within each Chinese province or EU member state). Although there are considerable challenges to such comparative analysis, it offers the possibility of enhancing understanding of migration in two very large economic units both of which are undergoing processes of transformation.

The paper starts by outlining the rationale for undertaking international comparative migration studies in general and for comparing China and the EU in particular. It then presents an initial exploration of three key areas: data issues in comparing the two geographical units; development and testing of migration flow models focusing on gross migration flows using fixed-effect negative binomial regression methods; and lastly the initial outputs of the models which are discussed in the context of developing future research directions.

2. International comparative migration analysis – contextualising comparing China and the EU

As Bell et al. (2002) suggest, international comparative migration has received relatively little attention, though recent years have seen a growth in such studies (see Bell et al. 2002: 435-6; Boyle et al. 2001; Skrentny et al. 2007; Bell and Rees 2006; Newbold and Bell 2001; Stillwell et al. 2001; Østbye and Westerlund 2007; Takenaka and Pren 2010; Yano et al. 2003; Sanderson 2010; Long et al. 1988; Kovacheva et al. 2012). This literature has compared a range of countries (Table 1), in some cases geographically proximate and relatively similar (e.g. Norway and Sweden) but others located far apart and with diverse geographies (e.g. the UK and Japan or Australia). Data are typically drawn from censuses, though other sources are utilised, and there is wide variety in the method and type of modelling. Finally, as Table 1 indicates there is considerable variation in the questions and processes which are explored and the indicators used in modelling. A coherent and systematic approach to international comparison has yet to be developed, suggesting that further studies should provide important case studies and tests of methods and theory to support such a development.

[Table 1]

One reason for the lack of studies is the difficulty of comparison due to the different characteristics and migration dynamics of countries, data format and availability, the questions asked and the models employed. However, the literature suggests that despite this international comparisons are a valid approach for migration studies to follow (Bloemraad 2013). Bell et al. (2002) suggest international comparison can enhance theorising about migration and policy recommendations as it can develop greater rigour in migration studies within and between countries and make national results more significant when placed in an international context.

Nevertheless, attempting to compare China and the EU may appear unrealistic. It could be argued, for example, that comparing a state with a supra-national organisation of several states involves comparing two geographical entities which are just too different, particularly in terms of their internal structures and regulation of migration. Furthermore, it could be argued that the two territories differ too widely with respect to their socio-economic and political characteristics.

However, we would argue that there are sufficient grounds for undertaking this comparison. An explicit East-West international comparison may yield important insights for understandings of globalisation, regionalisation and migration theory (and see Kovacheva et al. (2012) which suggests that such comparison could be fruitful). China is such a large country (in terms of its territory, population and economy) that finding other single states to compare it with is difficult. The logic of the development of the EU has been to create a larger political-economic entity that can compete with the world's biggest and growing economies, particularly China itself. Thus the EU is a geographical unit which could be considered large enough for comparison. Successive waves of enlargement have produced an EU comprising of twenty-seven states while China has twenty-two provinces, five autonomous regions, four municipalities under central administration and two special administrative regions (Hong Kong and Macau) (Figures 1 and 2). China has a population of 1.34 billion compared to the EU's 0.5 billion and an area of 3,705,386 square miles compared to the EU's 1,707,787 square miles in 2012 (the latter is likely to undergo further enlargement from 2013). China's economic size of $11.3 billion is similar to (and moving towards) the EU's $15.8 billion (purchasing power parity, 2011). In 2011 China recorded massive labour migration with 252.78 million migrants. Flows within the EU, particularly from the ten new member states into EU-15 countries, have grown since the enlargements of 2004 and 2007, although they do not match Chinese levels of mobility, averaging 250,000 citizens per year between 2004-9.

[Figure 1]

[**Figure 2**]

Both China and the EU are characterised by some administrative and political restrictions on mobility but in both cases these restrictions are being reformed to allow freer (although not completely free) movement. The EU internal market is based around the free movement of goods, services, capital and people - the latter a major goal of European integration from the 1950s. There are no restrictions on mobility of persons or households between most member states and migrants are accorded similar socio-economic status and rights. The removal of restrictions for movement between states within the EU has generated significant migration flows as it enlarged from the fifteen 'old' pre-1995 member states to twenty-five states in 2004 and twenty-seven in 2007, but is not complete or uniform. Five EU states (the UK, Cyprus, Bulgaria, Romania and Ireland) remain outside the Schengen Area, and some labour market migration restrictions were placed on some of the 2004 and 2007 Accession states, though these were phased out by 2011. However, the general trend has been a reduction of restrictions on the movement of citizens within the EU.

Inter-provincial migration in China continues to be regulated by the state, but there has also been a reduction in these restrictions in line with internal reforms and an incremental liberalisation of the economy. One characteristic which continues to limit internal mobility within China is the *Hukou* system (Chan 2010). *Hukou* has its origins in ancient China as a household registration system officially identifying the members of households as residents of an area. From the 1950s, the *Hukou* system categorized citizens according to place of residence and eligibility for certain socio-economic benefits. By the 1970s the system had become so rigid that reforms were necessary and as these accelerated during the late 1990s national and local authorities relaxed restrictions on obtaining urban residence permits. Nevertheless, *Hukou* still enforces restrictions based on income and housing criteria that work against rural-urban migration. For example, migrants who do not meet the criteria usually cannot obtain public services such as health care and schooling on an equal basis with other residents. Although China has exhibited large-scale inter-provincial migration in the last three decades these migration restrictions have had a significant effect on the scale and structure of migration (Fan 2005a; Bao et al. 2011; Whalley and Zhang 2007).

Given these characteristics comparisons between China and the EU do present problems but none necessarily more marked than in other attempts at international comparative analyses. Both are large-scale territorial units which can be said to exhibit comparable forms of inter-migration despite their different political-territorial organisation and structure. Both are major economic blocs within the global economy which are undergoing transition processes particularly with respect to labour migration. As such, attempts at comparison are valid and require development.

3. Data, modelling and determinants for comparing the EU and China

International comparative studies of migration are hampered by different national practices of data definition, measurement and availability (Long et al. 1988; Bell et al. 2002; Bell 2003). Bell et al. (2002) identify four key areas in which problems of comparison arise: measurement and types of data; temporal comparability; coverage of populations; and spatial units. In this section we therefore explore which data sets and variables can be compared for studies of inter-migration within the EU and China and identify which variables can be used in modelling.

In the EU, Eurostat is responsible for the collection and dissemination of migration data. A 2007 Regulation established common rules for the compilation of migration statistics in member states, in theory producing consistent data collection across the EU. However, Eurostat is limited to compiling aggregated data from national statistical offices on a relatively small number of topics (Kraler and Reichel 2010). The Labour Force Survey (LFS), coordinated by Eurostat, is in practice usually implemented differently in each member state. Overall, however, Martí and Ródenas (2007) concluded that estimates of stocks of immigrants are more robust than those of flows when using LFS data.

In China, the National Bureau of Statistics (NBS) is the organization authorised with collating population census data. There have been six national censuses since 1982 combined with an inter-census one per cent sample survey, most recently from 2005. Census data are aggregated to various levels of administrative units from town/township up to province. However, the definition of migrants has changed over the past five decades and there have been increases in spatial and temporal resolutions.

EU migration data requires standardisation to be comparable with China, particularly in relation to national differences in definitions (e.g. birth place or citizenship) and sources (Pedersen et al. 2008; Zaiceva and Zimmerman 2008; Kupiszewski and Kupiszewska 2010). The MIMOSA project (*Modelling of statistical data on migration and migrant populations* - http://mimosa.gedap.be) aimed to produce consistent estimates of international migration flows by country of origin and destination (Raymer and Abel 2008; De Beer et al. 2010; Abel 2010; Raymer et al. 2011). Long-term international migrants are defined as persons who move to a country other than their usual residence for at least one year. The MIMOSA depository includes estimates for 2002-7 of the migration matrix of flows by origin/destination (De Beer et al. 2010) which are relatively consistent across the EU.

It is also necessary to consider temporal comparability, population coverage and which spatial and other units are used. Table 2 thus summarises the approaches adopted in the literature analysing migration into and within the EU and China. This allows the identification of variables which can be compared for similar years and whose significance and influence has been demonstrated in previous models.

[**Table 2**]

Table 2 indicates considerable diversity in studies of inter-migration in the two contexts. Models of inter-provincial migration in China have a more consistent definition of migrants and identical spatial scale but lower temporal resolutions of migration flow data than models of international migration into EU and OECD countries. The definition of a migrant in China is dependent on the spatial and temporal scale used in the three population surveys (Fan 2005a). Sources of data utilised for migration study at the EU/OECD level are more diverse than those for China where data sets are largely dependent upon national censuses.

Some key variables suffer from inconsistent definition. For example, the geographic variable 'distance' is more commonly measured as the rail or road distance between provincial capitals in China (eg. Fan 2005; Cai and Wang 2003; Poncet 2006; Bao et al. 2011) but EU studies tend to use the direct line distance. However, despite this limitation distance remains a key determinant on migration and is therefore incorporated into the model developed here.

There are also diverse choices of modelling methods and dependent variables, which are affected by the availability of flow data, temporal scale and focus of study. For example, a number of studies of the EU and China (Table 2) point to the validity of incorporating migration flow as the dependent variable. Determinants models analysing China are dominated by the use of ordinary least squares (OLS) regression due to the lack of dynamic (e.g. annual) data.

Table 2 demonstrates that many of the determinants listed are significant for migration patterns in both areas, although the magnitude and sign of their influence may differ in each case. Eight explanatory variables are shown to have been statistically confirmed as significant determinants of internal migration in both contexts: population size, distance, contiguity, GDP per capita, income difference, unemployment rate, migration network, and migration restriction. These variables are therefore incorporated into the modelling (Table 3). However, it should be noted that migration is a complicated political, social, economic and environmental process and more demographic, social, economic, cultural, labour market, housing, environmental and regional variables should be incorporated into models in order to seek a full set of determinants (Fotheringham et al. 2004). Among the eight variables, population size, distance, income difference and network variables demonstrate consistent or robust contributions in China and the EU - either positive (+) or negative (-) - but the rest show inconsistent though significant contributions i.e. positive in some cases but negative in others. The policy variable 'migration restriction' has only relatively recently been confirmed as an important determinant and the literature generally lacks studies of its impact on migration (Bao et al. 2011). However, studies in both areas reveal that the incorporation of a migration restriction variable – transitional migration arrangements in the EU and *Hukou* in China - into models improves the assessment of internal mobility (Bao et al. 2011; Marques 2010; Poncet 2006).

Other variables whose significance varies between the EU and China are not incorporated into the modelling. For example, language (Lewer and Van den Berg 2008; Hooghe et al. 2008; Pedersen et al. 2008; Marques 2010), colonial history (Lewer and Van den Berg 2008; Hooghe et al. 2008; Warin and Svaton 2008) and social protection expenditure (Warin and Svaton 2008) have been confirmed as significant determinants in the EU/OECD. In contrast, foreign direct investment (FDI) (Chen and Coulson 2002; Cai and Wang 2003; Bao et al. 2011) and consumption level (Cai and Wang 2003) have been claimed as significant determinants in China.

Finally, currently it is only realistic to compare inter-state migration in the EU for 2007 with inter-province migration in China for 2005. Romania and Bulgaria joined the EU in 2007 and 27 member states are numerically closer to China's thirty-two provincial units, and the most recent matrix of flow data for China is from the 2005 inter-census, making these the most suitable years for comparison. However, it should be noted that the selection of a comparable year on both sides is not straightforward as immigration policies across the EU have changed, compared to the relatively established system in China. Additionally, EU enlargement in 2004 and 2007 produced new (and in some cases significant) flows, whereas Chinese flows have been more stable. As a result, a panel data set over a long period should be collected in the future to reflect the significant influence of EU enlargement on inter-state mobility. In the following section we describe the development of migration flow models using these variables.

4. Developing migration flow models for China-EU comparisons

This section will focus on measuring patterns of inter-migration and modelling the determinants of inter-provincial migration in China and inter-state migration within the EU using comparable macro-statistical data sets.

4.1 Measuring patterns of migration flows

Bell et al. (2002) proposed four groups of measures, each of which provides a different insight into migration: intensity, distance, connectivity and the effect of migration. These measures have been employed for comparisons of Britain and Australia, for example (and see Stillwell et al. 2001; Stillwell and Hussain 2010), and help analyse the inequality, connectivity, effectiveness, and distance decay effect of migrants' flows and provide objective and robust quantitative indicators for comparing flow patterns. Here we begin to describe the modelling approach developed for the comparison.

To begin with, M_{ij} denotes the number of migrants from origin unit i to destination unit j, d_{ij} the distance between the centroids of the two units, and n the total number of units. The average distance (D') is measured by the equation (1):

$$D' = \frac{\sum_i \sum_j M_{ij} * d_{ij}}{\sum_i \sum_j M_{ij}} \qquad (1)$$

The spatial interaction model of migrant flow is represented as equation (2):

$$M_{ij} = A_i * B_j * O_i * D_j * d_{ij}^{-b} \qquad (2)$$

where O_i is the total of out-migrants from unit i to all destination units and D_j the total of in-migrants to unit j from all origin units. A_i and B_j are balancing factors. The distance decay parameter *b* will be calibrated using a negative binomial regression due to the over-dispersion of M_{ij} (see details below).

The migration connectivity (MC) is represented as equation (3):

Define C_{ij}=1 if $M_{ij} > M_0$ otherwise C_{ij}=0 then:

$$MC = \frac{\sum_i \sum_j C_{ij}}{n*(n-1)} \tag{3}$$

This indicator ranges between 0 and 1, depending on the definition of a threshold value of gross flow M_0. MC is the simplest measure reflecting the strength of functional linkage or interaction between the predefined spatial units.

The index of migration inequality (IN) is represented as equations (4) and (5):

$$IN = 0.5 * \frac{\sum_i \sum_j abs(M_{ij} - M'_{ij})}{\sum_i \sum_j M_{ij}} \tag{4}$$

$$M'_{ij} = \frac{\sum_i \sum_j M_{ij}}{n*(n-1)} \tag{5}$$

M'_{ij} indicates the expected distribution of migrant flows (measured as the mean value of all flows). The index of migration inequality (IN), ranging from 0 to 1, reflects the difference between the observed distribution and expected distribution of gross flows. A higher value suggests a more uniform pattern and a smaller value indicates greater inequality.

The coefficient of variation (CV) is measured as the standard deviation divided by the mean migrant flow. CV is an interpretable index measuring spatial focusing of migration (Fan 2005a). A greater value suggests a higher concentration of flows.

Migration dispersion (DI) can also be measured based on entropy (EN) which is commonly used for quantifying diversity:

$$DI = \frac{(Log(n^2) - EN)}{Log(n^2)} \tag{6}$$

$$EN = \sum_i \sum_j (\frac{M_{ij}}{S}) * Log(\frac{S}{M_{ij}}) \tag{7}$$

$$S = \sum_i \sum_j M_{ij} \tag{8}$$

Equation (6) aims to normalise the entropy values for the purpose of relative measurement. DI close to zero indicates high dispersion in migration flow distribution, while DI close to one is the maximum concentration.

The effectiveness index (EI) is calculated as in equation (9):

$$EI = 100 * \frac{\sum_i abs(D_i - O_i)}{\sum_i (D_i + O_i)} \tag{9}$$

where O_i is the total outflow from unit *i* and D_i is total inflow to unit *i, abs* is an arithmetic function of calculating its absolute value.

The EI index ranges from 0-100. Values close to 100 signify unidirectional movements from one unit to another, while values close to 0 indicate roughly equal flows in the two directions. Therefore migration effectiveness EI is useful for measuring the redistribution of migrants between spatial units.

[**Table 3**]

4.2 Modelling determinants of migration

4.2.1 *Variables*. There has been considerable debate regarding the definition of the dependent variable in migration models (*net*, *gross* or *rate of* migration). Where out- and in-flows are correlated net migration has resulted in biased estimates of variable coefficients and thus it is useful to estimate a model using gross migration flow or rate as dependent variables (Van der Erf and Heering 2002). Selecting M_{ij} as the dependent variable is relevant given the focus of this research. Figures 3 and 4 reveal the spatial distributions of gross migrant flow across the EU (2007) and China (2005) respectively.

Table 3 summarizes the definitions of, and data sources for, the variables created. All numerical explanatory variables (except the binary variable 'Contiguity') are log-transformed to meet the requirement of normal distribution. Among these explanatory variables, political, economic and demographic variables have either one or two years lag, depending on the availability of data sets. The network variable is five-year lagged to control for the possibility of endogeneity.

4.2.2 *Models*. A comprehensive and comparable equation of inter-migration must incorporate all the variables defined above, thus:

$M_{ij} = P_i^{\beta 1} * P_j^{\beta 2} * D_{i,j}^{\lambda} * N_{i,j}^{\beta 3} * T_j^{\beta 4} * G_{j,i}^{\beta 5} * E_{j,i}^{\beta 6} * W_{j,i}^{\beta 7} * C_{i,j}^{\beta 8} * \exp\varepsilon_r$
(10)

where $\beta_1 ... \beta_8$ and λ are the respective coefficients and ε_r is an error term.

Equation 10 requires further refinement. There are two main regression options for calibrating Equation 10 - Ordinary Least Squares (OLS) regression and Poisson regression.

A linear calibration of Equation 10 by OLS (with logarithms taken on both sides) yields:

$LnMij = LnK + \beta_1 * LnP_i + \beta_2 * LnP_j + \lambda * \mathrm{Ln}D_{i,j} + \beta_3 * LnN_{i,j} + \beta_4 * LnT_j + \beta_5 * LnG_{j,I} + \beta_6 * LnE_{j,I} + \beta_7 * LnW_{j,I} + \beta_8 * C_{i,j} + \varepsilon_r$ (11)

where K is a constant.

OLS calibration assumes that the dependent variable is normally distributed but flows of migration (count) data are highly non-normal and impossible to log transform when flows (M_{ij}) are zero. Therefore a Poisson formulation of the spatial interaction model is:

$M_{ij} = \exp(K + \beta_1 * LnP_i + \beta_2 * LnP_j + \lambda * \mathrm{Ln}D_{i,j} + \beta_3 * LnN_{i,j} + \beta_4 * LnT_j + \beta_5 * LnG_{j,I} + \beta_6 * LnE_{j,I} + \beta_7 * LnW_{j,I} + \beta_8 * C_{i,j}) + \varepsilon_r$ (12)

The small number of zero flows is not problematic for the Poisson formulation which considers migration as a stochastic process in which events (migration flows) occur continuously and independently of one another. It assumes that the dependent variable is neither over-dispersed nor has an excessive number of zeros. Over-dispersion means that the conditional mean of the dependent variable (flow) is significantly different from the conditional variance. When there is over-dispersion, the estimates are consistent, but inefficient. To solve the issue of over dispersion we apply negative binomial regression.

Using migration flows without time series data may have un-observed spatial heterogeneity, meaning different migration behaviour among the spatial units, such as varied selectivity of destination units between origin units. In general, spatial heterogeneity measures such kinds of spatial variation across a study area. Spatial heterogeneity in this case may be demonstrated by the spatial variation in intercept (δ_i in Equation 13) of regression equations, which is called a fixed effect model. Fixed effect in terms of spatial units can be modelled at origin level, destination level or even origin-destination level. In this paper, we focus on origin-specific effects.

A fixed effect model aiming to explore spatial heterogeneity at origin unit level is thus:

$M_{ij} = \exp(K + \beta_1 * LnP_i + \beta_2 * LnP_j + \lambda * \mathrm{Ln}D_{i,j} + \beta_3 * LnN_{i,j} + \beta_4 * LnT_j + \beta_5 * LnG_{j,I} + \beta_6 * LnE_{j,I} + \beta_7 * LnW_{j,I} + \beta_8 * C_{i,j} + \delta_i)$ (13)

where δ_i is the intercept estimated for the *ith* origin spatial unit, representing the origin unit's contribution to the spatial variation of migration flows across the study area fixed effect.

Allison and Waterman (2002) stated that the conditional fixed-effect negative binomial model is not a true fixed-effect method as it does not control for all stable covariates. Their simulation study yields good results from applying an unconditional negative binomial regression estimator with dummy variables to represent the fixed effects, but the standard error estimates should be corrected using the deviance statistic which can be implemented by the robust standard error in the *Stata* software package (Guimarães 2008). This method was used to calibrate Equation 13 (following Marques (2010)).

4.3 Spatial patterns of intermigration flows

4.3.1 *Comparisons of mobility patterns between the EU and China*.

Model outcomes are presented here to explore, first, if any meaningful results can be produced through such a comparison and, second, as the basis for developing further comparative research.

Table 4 provides a statistical summary of flow patterns (M_{ij}). Both flows exhibit a high degree of over-dispersion as their variance is much greater than their mean and there are many observations with zero flow, resulting in the selection of negative binomial regression for these models.

[**Table 4**]

[**Figure 3**]

[**Figure 4**]

[**Figure 5**]

Figures 5 and 6 reveal the patterns of migration inflows and outflows across member states and provinces. In the case of the EU, both inflow and outflow migration was dominated by the UK, Germany, France, Spain, Poland and Romania. A high level of mobility exists in both western and eastern regions. Figure 5 demonstrates a very high correlation between inflow and outflow (correlation coefficient = 0.667). In the case of China (Figure 6), outflow migration is dominant in the central region and inflow migration more significant in eastern areas. Compared to the EU, there is a very low correlation (coefficient = 0.095) between out- and in-flows in China. There are many zero flows (M_{ij}) on both sides thus models should focus on gross migration instead of net migration and Stata 11.0 software was therefore used to build negative binomial standard and fixed effect models at the defined levels.

[**Figure 6**]

[**Table 5**]

4.3.1 *Measurement of mobility pattern*.

In Table 5, the values for average distance and the distance decay parameters reveal that Chinese migrants travelled longer distances and at the same time were more sensitive to distance than the EU27 migrants. Longer average distance in China is the result of China's vast size. However, more sensitivity to distance points to the increased significance of physical distance in selecting a destination province. Here modelling requires a more sophisticated measure of distance, as what this result probably reflects is the impact of relative travel costs. Multiple trips between home and work place by train or coach by migrants and their families is still a negative factor in China (as is confirmed in the models later).

The high connectivity (MC) scores indicate very strong connections between all spatial units in both regions in terms of direction but not magnitude. The score for inequality for China (0.711) suggests a much greater tendency towards inequality than the EU27 though both showed evidence of inequality. This has been further confirmed by the values for the co-efficient of variation. The values of migration effectiveness (E_i) suggest that migration appears to be more efficient as a mechanism for redistributing migrants between province level units in China than between member states in the EU. The flows of migrants are dominated by unidirectional movements from one province to another in China.

4.4. Model results and findings

[**Table 6**]

Table 6 lists the summary statistics of the non-categorical variables. Multicollinearity within the data presents problems. In the EU models, a high correlation exists between the dependent variable and stock of migrants (co-efficient = 0.3755) and between the ratio of GDP per capita and the ratio of average wages (0.912). In the China model, high correlation is present between the ratio of GRP per capita and the ratio of average wage (0.366) and between the ratio of unemployment rates and the ratio of average wages (0.36). Accordingly, the variable ratio of average wages was removed from modelling on both sides.

The initial gravity model considered two variables: population size in origin and destination units and the distance between origin and destination units. Table 7 lists the results of the negative binomial regression models. Distance is considered a fundamental explanatory variable which proxies migration costs (Greenwood 1985; Greenwood and Hunt 2003). Alpha is the estimate of the dispersion parameter (data are over dispersed when alpha is significantly > zero). Alpha coefficients (3.38 for China and 1.67 for the EU) confirm that inter-provincial mobility in China is much more dispersed than in the EU. This might be caused by the marked spatial inequality of economic development across China between economically rich Eastern regions and deprived Western regions. The elasticity of distance (-1.320609 compared to -0.7617719) confirms that inter-province mobility in China is more sensitive to travel distance than in the EU. One factor here might be that inter-provincial migration in China has been dominated by rural-to-urban migration and rural people are very concerned about the costs of travel in selecting a destination city. However, inter-state migration across the EU is more dominated by urban-to-urban migration and here cultural distance (e.g. language and colonial history) may be more important than physical distance (Lewer and Van den Berg 2008; Hooghe et al. 2008; Pedersen et al. 2008; Marques 2010; Warin and Svaton 2008).

Origin population size demonstrates a positive contribution to the flow of migrants in both cases, but destination population size shows contrasting effects: positive for the EU but negative for China. This shows that population size at origin sites is still a remarkable push factor across both contexts. Population size at destination sites is only a significant pull factor in the EU as population size is highly correlated with economic power at member state level. For example, the UK and Germany are the greatest receivers of migrants, against the Pearl and Yangtze Deltas in China.

[**Table 7**]

[**Table 8**]

[**Figure 7**]

[**Figure 8**]

Table 8 lists the outcomes of the two full models for the EU and China taking all independent variables into account. The fixed effect model for the EU demonstrates that all explanatory variables - except the transitional arrangements - show statistically significant contributions to the flow of gross migrants. Among the five significant variables, spatial adjacency, ratio of unemployment rates and network effect all make positive contributions, in contrast to two other variables - distance and the ratio of GDP per capita - which make negative contributions. The insignificance of transitional arrangements further confirms the conclusions of Zaiceva and Zimmermann (2008) that there is no conclusive evidence for a direct link between the scale of migration and transitional arrangements.

The fixed effect model for China demonstrates that all explanatory variables show statistically significant contributions to the flow of gross migrants. With the exception of the distance variable, all of them exhibit a positive contribution to the flow of migrants.

Some initial comparisons can be made based on the model outcomes (Table 8). First, distance is a negative explanatory variable in both cases but elasticity is much higher in the case of China (-.498) than in the EU (-.231), meaning inter-provincial migration in China is more sensitive to travel distance (confirming the conclusion drawn from Table 7). This is possibly due to it being more common for family groups to migrate in China which greatly increases transportation costs compared to single migrants.

Second, the ratio of pcGDP and ratio of unemployment rates show contrasting effects. The ratio of pcGDP is negative in the EU but positive in China. The ratio of unemployment rates exhibits a strongly positive influence in the EU but a relatively weaker significance in China. This indicates that the unemployment rate is one of the key pull factors attracting migrants in the EU but that the ratio of GDP per capita is one of the dominant pull factors in China. Or rather, EU migrants travel across member states for jobs but Chinese migrants travel across provinces for higher income.

Third, network effect is a significant positive factor in both cases but it has a stronger impact on flows of migrants in the EU. This raises the question of whether family ties in the case of China and membership of the same national or ethnic group in the case of EU help reduce the economic and social costs of migration. Migration network is thus further confirmed here as a key determinant of migration in two very contrasting socio-economic and cultural and political contexts.

Finally, migration restriction policy is not significant in the EU in contrast to the positive contribution of Chinese *Hukou*. Again, it is difficult to make a straightforward comparison as EU migration policy demonstrated both restrictive and liberal approaches in 2001-06 (Marques 2010). This can partly explain the patterns of gross migrant flows, particularly why the flows in China are much more dispersed than those across the EU as *Hukou* policy makes a significant contribution to this pattern.

Both fixed-effect models also reveal that there is un-observed spatial heterogeneity present with the flows of gross migrants across the EU and China. Spatial heterogeneity is represented by the intercept coefficients of administrative units (see Figures 7 (the EU) and 8 (China)). Figure 7 indicates that the UK, Germany, Spain, Poland and Romania made greater contributions to the flows of gross migrants than the rest, contrasting with the least contributing members - Norway, Finland, Italy and Belgium. Figure 8 demonstrates that Sichuan, Chongqing, Henan and Zhejiang are the most active provinces, contrasting with those least active – Shaanxi, Qinghai, Tibet, Hainan, Tianjin and Shanxi. The eastern region is much more active than the western region and sixty-five per cent of inter-provincial migrants arrived in the Eastern region in 2000 (Cai and Wang 2003).

5. Conclusions and research directions

This paper argues that comparing inter-migration in the EU and China is both necessary and feasible, although the paucity of theoretical frameworks, appropriate methods and comparable data sets remains an issue. The analysis presented here represents a first attempt at tackling such a comparison. The purpose of the paper has been to explore the issues involved, evaluate data sources and undertake some initial modelling to investigate whether realistic and meaningful comparisons can be made.

Flow patterns and determinants of inter-migration have been successfully quantitatively explored laying the foundation for future research including the development of theoretical frameworks and methodological and technical solutions. The overall conclusion of this paper is that such comparisons are a realistic and useful topic for analysis and we conclude by exploring some of the potential key questions, implications for theory and methodological challenges.

Some of the key theoretical concerns of migration which are important for gravity modelling and neoclassical approaches are confirmed here, namely the importance of space in terms of distance and contiguity. The comparison reveals a difference between China and the EU, however, with distance apparently a stronger influencing factor in the former. The actual influence of distance therefore requires further investigation. In China sensitivity to distance seems to be higher and this probably relates to the issue of the costs of mobility. Travel costs within China remain relatively high compared to within the EU. If migration flows in China are more dominated by family units then this will intensify this effect. This also has implications for modelling the influence of distance as current measures of straight-line distances or distances between major centres (which models currently use) may not account for the relative costs of covering distance for migrants.

Considerations of distance must also become more nuanced in their appreciation of socio-cultural differences, which critics note have been largely ignored in neoclassical approaches to theorizing migration, particularly the costs of cultural adaptation. In China this may be less of an issue as language and socio-cultural practices are more similar and the costs of cultural adaptation lower, but sensitivity to physical distance seems to be higher. Therefore, an important question is whether socio-cultural distance is more significant in the EU in determining migration. One example might be seen in the propensity for Polish migrants to favour the UK for historical reasons while Romanian migrants have favoured Italy and Spain because of similarities in language, climate and culture.

The role of economic determinants has been debated in a number of theoretical approaches. Results from the initial model presented above suggest that in the Chinese case wage rates are an important determinant but in the EU it is the availability of jobs which is more important. Neoclassical theory on migration argues that disparities in wage rates are important and that migrants are rational actors who will move to maximise the rewards of their labour ie. wages. Here, the Chinese case offers some support for such a view. This also links to the new economics of labour migration theory (Stark and Taylor 1989; Stark 1991) which emphasizes that the distribution of income or the relative deprivation of households in origin or receiving sites is important which again seems to be important in the Chinese case. However, the EU case would seem to offer a different perspective (job availability rather than wages).

The reasons for this require more exploration. It would also be important to examine what the comparison reveals for dual labour market theory (Piore 1979), which posits that migrants move internationally because there is demand for foreign labour to fill low-wage, low-skill, unpleasant jobs. Is this explanatory framework applicable at all in China, resting as it does on international movement of foreign labour? Or, does China exhibit some form of segmented labour markets which are important for driving migration? Overall, however, the results suggest a continued role for neo-classical ideas about the role of spatial inequality or the imbalance of economic development, as difference of income or wage is a strong push factor. In particular, the new economics of labour migration theory may benefit from examination of this distinction. This theory points to the family (rather than the individual) as the key actor seeking to enhance its utility, and this seems to be more applicable in the Chinese case, where migration often involves family units rather than single migrants. The distribution of income and the relative deprivation of households in origin or receiving sites is an important driving force in China.

Importantly, the comparative analysis highlights the importance of migration network theory (Massey et al. 1998), as network variables have a significant influence on flows of migrants in both contexts, supporting the suggestion that networks rank among the most important explanatory factors of migration (Arango 2000). What requires further examination here, however, is the different characteristics of networks and how they operate in both contexts. Is the EU as characterised by younger, single migrants while Chinese mobility is more typified by family units moving for work opportunities? And what about the relative propensity to return migration, which was considered an important likely phenomenon in the case of some EU migrant flows (particularly flows from the 2004 and 2007 accession countries) but which may also be significant in China as there are advantages in retaining a rural *Hukou* registration? Neoclassical approaches have also been criticised for overlooking social factors and within the flows which have been identified further research could distinguish their composition and characteristics in terms of age, educational background and family circumstances.

Finally, another implication from this comparison matches criticisms of the neo-classical approach to explaining migration that it fails to accord much of a role to political factors (instead prioritising factors such as wage rates). An EU-China comparison opens up a key opportunity to explore the influence of political variables which influence migration in more depth as well as the political processes which have shaped mobility patterns in the two areas in the past, factors which are becoming increasingly recognised in migration theory. As Arango (2000: 293) suggests 'the powerful impact of admissions restrictions on processes, determinants and selectivity should be incorporated as an essential ingredient in models.' Political factors such as immigration policy on the EU side and *Hukou* regulation in China have varied impacts on flows. There are much more complicated mobility related policies across the EU than in China where *Hukou* is a solo dominant factor. Here comparative modelling revealed that *Hukou* restriction is a more significant factor than EU policy, but the actual reasons for this difference will require finer-detailed studies to evaluate this finding.

Answering these questions requires greater methodological sophistication and the challenges relating to data and models must receive substantial attention. The pattern analysis (including exploratory indicators and mapping) presented here is based on a flow data matrix of migration. This analysis is visual and summative so it is highly interpretable and useful for comparative studies, but its success is reliant on the quality of data as measurement errors in explanatory variables such as wage rate in the China side would make regression models less efficient and the coefficients produced might be biased towards zero (Aydemir and Borjas, 2011). The harmonization of EU flow data has remarkably improved the pattern analysis and made it more comparable with China. Considering the varied definitions and data sources of other socio-economic, political and environmental determinants of migration across the EU member states such harmonization should be extended to cover all other relevant data sets. Flow data is very informative for representing and understanding spatial interaction but current methods of pattern analysis, particularly visualisation (e.g. 2D mapping), are still weak for comparing spatial interaction. More time series flow data and more advanced methods such as spatial dynamic panel models (Mitchell et al. 2011) should be developed in the future, which can take account of spatial dependence and heterogeneity in models.

The types of models which can be used to explore the significance of the determinants of inter-provincial or inter-state mobility are highly dependent on the format of data which is available in the two areas to be compared. Macro-models usually employ cross sectional data from census surveys. It is clear that the success of these macro models is very much dependent on the quality of available data and their comparability between countries. As a first exploration, only a few variables which had previously been proven to be significant determinants of mobility in both the EU and China were chosen for comparison. An important future development will be the identification of a greater range of variables which could be included into comparative models and the associated development of standardised datasets. In contrast, micro-models considering socio-cultural factors more typically utilise data from sample surveys at the individual level (e.g. a person or household). These offer the potential to overcome some of the data issues that census surveys face and might be complementary to macro-models in terms of identifying determinants (factors influencing a migrant decision) and developing data comparability (see the example in Takenaka and Pren (2010)). This represents a huge challenge for research as the collation of a statistically meaningful sample size in both the EU and China would be a substantial task.

The modelling exercise presented here has demonstrated that fixed-effect models can explain the unobserved spatial heterogeneity present in spatial units. However, the spatial dependence, or rather the spatial autocorrelation in the residuals of these models such as negative binominal regression models, has not yet been considered, which involves the calculation of spatial autocorrelation of interaction or flow data. This will be a key aspect in the development of future flow models. It will also be important to explore and determine which spatial and temporal scales are the optimum for comparing inter-regional migration between the two territories. Intra-migration (within member states in the case of EU and within province units or even county units in the case of China) is also substantial on both sides. This comparison may reveal different findings and provide varied insights into migration pattern and process and this topic should be further studied in the future.

A further aspect of modelling will be to develop dynamic datasets which will support temporal analyses. In this initial analysis it was only possible to compare one chosen period - 2007 in the EU and 2005 in China. However, both territories have a long history of inter-migration and changes in the regulation of that migration, so an overview of historical changes in patterns would deepen understanding of this complicated political and socio-economic process. Towards this, further analyses and models using panel data (time series, cross-sectional data) from both geographical units would facilitate dynamic comparisons, which could reveal and help explain similarities and differences in mobility over a longer-term political and social context. The challenges for such dynamic comparisons include the processing (e.g. harmonization, smoothing and standardization) of dynamic data and development of dynamic models.

Finally, such challenging comparative studies will require international and multi-disciplinary collaboration between the EU and China. The formation of international research teams, with governmental support, capable of undertaking these advances represents a significant logistical and financial challenge in the development of international comparative migration studies. Such comparative studies would facilitate the development of middle-range theories (Castles, 2010) by providing diverse insights into varied patterns and processes of migration.

Conflict of interest statement: None.

References

Abel, G. J. (2010) ‘Estimation of international migration flow tables in Europe’, *Journal of the Royal Statistical Society Series A*, 173/4: 797-825.

Allison, P. D. and Waterman, R. P. (2002) ‘Fixed-effects negative binomial regression models’, *Sociological Methodology*, 32: 247-65.

Arango, J. (2000) ‘Explaining migration: a critical review’, *International Social Science Journal*, 165: 283-95.

Aydemir, A., Borjas, G. J. (2011) ‘Attenuation bias in measuring the wage impact of immigration’, *Journal of Labor Economics*, 29/1: 1-45.

Bao, S., Bodvarsson, Ö. B., Hou, J. W. and Zhao, Y. (2011) ‘The regulation of migration in a transition economy: China’s Hukou system’, *Contemporary Economic Policy*, 29/4: 564–79.

Beine, M., Docquier, F. and Rapoport, H. (2007) ‘Measuring international skilled migration: a new database controlling for age of entry’, *World Bank Economic Review*, 21/2: 249–54.

Bell, M., Blake, M., Boyle, P., Duke-Williams, O., Rees, P., Stillwell, J. and Hugo, G. (2002) ‘Cross-national comparison of internal migration: issues and measures’, *Journal of Royal Statistical Society A*, 165/3: 435–64.

Bell, M. and Rees, P. (2006) ‘Comparing migration in Britain and Australia: harmonisation through use of age-time plans’, *Environment and Planning A*, 38: 959-88.

Bell, M. (2003) ‘Comparing internal migration between countries: measures, data sources and results’, *Paper Presented in Population Association of America 2003,* Minneapoplis, May 1-3.

Bloemraad, I. (2013) ‘The promise and pitfalls of comparative research design in the study of migration’, *Migration Studies*, 1/1: 27-46.

Boyle, P., Cooke, T. J., Halfacree, K. and Smith, D. (2001) ‘A Cross-National Comparison of the Impact of Family Migration on Women’s Employment’, *Demography*, 38/2: 201-13.

Castles, S. (2010) ‘Understanding global migration: a social transformation perspective’, *Journal of Ethnic and Migration Studies*, 36/10: 1565-86.

Cai, F. and Wang, D. (2003) ‘Migration as marketization: what can we learn from China’s 2000 census Data?’, *The China Review*, 3/2: 73-93.

Chan, K. W. (2010) ‘The household registration system and migrant labor in China: notes on a debate’, *Population and Development Review*, 35/2: 357-64.

Chen, A. and Coulson, N. E. (2002) ‘Determinants of urban migration: evidence from Chinese cities’, *Urban Studies*, 39/12: 2189-97.

De Beer, J., Raymer, J., Van de Erf, R. and van Wissen, L. (2010) ‘Overcoming the problems of inconsistent international migration data: a new method applied to flows in Europe’, *European Journal of Population*, 26/4: 459-81.

Fan, C. (2005a) ‘Interprovincial migration, population redistribution, and regional development in China: 1990 and 2000 census comparisons’, *The Professional Geographer*, 57/2: 295-311.

Fan, C. (2005b) ‘Modelling inter-provincial migration in China, 1985–2000’, *Eurasian Geography and Economics*, 46/3: 165-84.

Fotheringham, A. S., Rees, P., Champion, T., Kalogirou, S. and Tremayne, A. R. (2004) ‘The development of a migration model for England and Wales: overview and modelling out-migration’, *Environment and Planning A*, 36: 1633-72.

Greenwood, M. J. and Hunt, G. L. (2003) ‘The early history of migration research’, *International Regional Science Review,* 26: 3–37.

Greenwood, M. J. (1985) ‘Human migration: Theory, models and empirical studies’, *Journal of Regional Science*, 25: 521–44.

Guimarães, P. (2008) ‘The fixed effects negative binomial model revisited’, *Economics Letters*, 99: 63–6.

Hooghe, M., Trappers, A., Meuleman, B. and Reeskens, T. (2008) ‘Migration to European countries: a structural explanation of patterns, 1980–2004’, *International Migration Review,* 42/2: 476–504.

Kraler, A. and Reichel, D. (2010) *Statistics on Migration, Integration and Discrimination in Europe*. PROMINSTAT Final Report. Vienna: ICMPD.

Lewer, J. J. and Van den Berg, H. (2008) ‘A gravity model of immigration’, *Economics Letters*, 99: 164–67.

Long, L. Tucker, C.J. and Urton, W.L. (1988) ‘Migration distances: an international comparison’. *Demography,* 25:633–640.

Martí, M. and Ródenas, C. (2007) ‘Migration estimation based on the labour force survey: an EU-15’, *International Migration Review,* 41/1: 101–26.

Marques, H. (2010) ‘Migration creation and diversion in the European Union: is Central and Eastern Europe a ‘natural’ member of the single market for labour?’, *Journal of Common Market Studies*, 48/2: 265–91.

Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A. and Taylor, J. E. (1998) *Worlds in motion. Understanding international migration at the end of the millennium*. Oxford: Clarendon Press.

Mitchell, J., Pain, N. and Riley, R. (2011) ‘The drivers of international migration to the UK: a panel-based Bayesian model averaging approach’, *The Economic Journal*, 121: 1398–444.

Newbold, K. B. and Bell, M. (2001) ‘Return and onwards migration in Canada and Australia, evidence from fixed interval data’, *International Migration Review*, 35/4: 1157-84.

Østbye, S. and Westerlund, O. (2007) ‘Is Migration Important for Regional Convergence? Comparative Evidence for Norwegian and Swedish Counties, 1980-2000’, *Regional Studies*, 41/7: 901-15.

Pedersen, P. J., Pytlikova, M. and Smith, N. (2008) ‘Selection and network effects - migration flows into OECD countries 1990–2000’, *European Economic Review*, 52: 1160–86.

Piore, M. J. (1979) *Birds of passage: migrant labour in industrial societies*. Cambridge: Cambridge University Press.

Poncet, S. (2006) ‘Provincial migration dynamics in China: Borders, costs and economic motivations’, *Regional Science and Urban Economics*, 36: 385-98.

Raymer, J. and Abel, G. (2008) ‘The MIMOSA model for estimating international migration flows in the European Union’ Geneva, Switzerland, UNECE / Eurostat, 13pp. (UNECE/Eurostat work session on migration statistics Working papers, 8).

Raymer, J., de Beer, J. and van der Erf, R. (2011) ‘Putting the pieces of the puzzle together: age and sex-specific estimates of migration amongst countries in the EU/EFTA, 2002–2007’, *European Journal of Population*, 27: 185–215.

Sanderson, M. R. (2010) 'International migration and human development: a cross-national analysis of less-developed countries, 1970-2005'. *Social Indicators Research* 96(1): 59-83.

Shen, J. (2012) 'Changing patterns and determinants of interprovincial migration in China 1985–2000', *Population, Space and Place*, 18: 384–402.

Shen, J. (2013) 'Increasing internal migration in China from 1985 to 2005: Institutional versus economic drivers', *Habitat International*, 39: 1-7.

Skrentny, J. D., Chan, S., Fox, J. and Kim, D. (2007) 'Defining nations in Asia and Europe: a comparative analysis of ethnic return migration policy', *International Migration Review*, 41/4: 793-825.

Stark, O. (1991) *The migration of labour*. Cambridge: Blackwell.

Stark, O. and Taylor, J. E. (1989) 'Relative deprivation and international migration', *Demography*, 26/1: 1-14.

Stillwell, J., Bell, M., Blake, M., Duke-Williams, O. and Rees, P. (2000) 'A comparison of net migration flows and migration effectiveness in Australia and Britain: Part 1, total migration patterns', *Journal of Population Research*, 17/1: 17-38.

Stillwell, J., Bell, M., Blake, M., Duke-Williams, O. and Rees, P. (2001) 'Net migration flows and migration effectiveness: a comparison of Australia and the United Kingdom, 1976-1996. Part 2, age-related migration patterns', *Journal of Population Research*, 18/1: 19-39.

Stillwell, J. and Hussain, S. (2010) 'Exploring the ethnic dimension of internal migration in Great Britain using migration effectiveness and spatial connectivity', *Journal of Ethnic and Migration Studies*, 36/9: 1381-403.

Takenaka, A. and Pren, K. A. (2010) 'Determinants of Emigration: Comparing Migrants' Selectivity from Peru and Mexico', *Annals of the American Academy of Political and Social Science,* 630: 178-93.

Van der Erf, R. and Heering, L. (2002) *Moroccan migration dynamics: prospects for the future*. Geneva: International Organisation for Migration.

Kovacheva, V., Vogel, D., Zhang, X. and Jordan, B. (2012) 'Comparing the development of free movement and social citizenship for internal migrants in the European Union and China – converging trends?', *Citizenship Studies*, 16/3-4: 545-61.

Warin, T. and Svaton, P. (2008) 'European migration: welfare migration or economic migration?', *Global Economy Journal*, 8/3: 1-30.

Wu, Z. and Yao, S. (2003) 'Intermigration and intramigration in China: a theoretical and empirical analysis', *China Economic Review*, 14: 371-85.

Yano, K., Nakaya, T., Fotheringham, A. S., Openshaw, S. and Ishikawa, Y. (2003) 'A comparison of migration behaviour in Japan and Britain using spatial interaction models', *International Journal of Population Geography*, 9: 419–31.

Zaiceva, A. and Zimmermann, K. F. (2008) 'Scale, diversity, and determinants of labour migration in Europe', *Oxford Review of Economic Policy*, 24/3: 427–51.

Table 1. Comparative studies on migration related themes

Case studies	**Countries**	**Focus**	**Data**	**Methods**
Boyle et al. 2001	UK & USA	Impacts of migration (Probability of un-employability between male and female migrants)	Census micro-data	Logit model
Skrentny, et al. 2007	East Asia & Europe	Ethnic preferences	Qualitative document	Case studies
Bell and Rees, 2006	UK and Australia	Migration intensity	Census (Australia) and register data (UK)	Age-time diagrams, data harmonisation.
Newbold and Bell, 2001	Canada and Australia	Characteristics, patterns and spatial impacts of return and onwards migration	Censuses (1991)	Statistical summary
Stillwell, et al., 2000, 2001	UK and Australia	Migration effectiveness in relation to age groups and spatial distribution.	Census (Australia) and register data (UK)	Indicator, mapping
Østbye and Westerlund, 2007	Norway and Sweden	The effect of migration on economic convergence of regions	Diverse sources from each side	Dynamic panel data models (GMM system)
Takenaka and Pren, 2010)	Peru and Mexico	Migrant selectivity	Random sampling of households conducted by two projects	Event history analysis
Yano, et al. 2003	Britain and Japan	Migrants' behaviour	Census (1990 for Japan and 1991 for UK)	Spatial interaction model, Poisson regression (local), mapping
Kovacheva, et al., 2012	China and the EU	Social rights of migrants	No data used	Literature review

Table 2. A comparison of migration studies and determinants in the EU and China

Items	The EU	China
Definitions of Migrants	Citizenship (Hooghe et al., 2008; Marques, 2010); Mixture of citizenship, nationality and country of birth (Pedersen et al., 2008); Nationality (Warin and Svaton, 2008), Skilled migrants (Beine et al., 2007).	1990 census (moved from jiedao, town or township and has lived in this place for more than 6 months but less than 5 years); 2000 census and 2005 1% sample survey (more than 1 year but less than 5 years)
Spatial scale (origin/destination)	16 OECD countries (Lewer and Van den Berg, 2008); OECD countries (Beine, et al., 2007); 21 OECD-Europe countries (Hooghe et al., 2008); EU14 (Warin and Svaton, 2008); EU15 (Marques, 2010)	Provincial units (Cai and Wang, 2003; Fan, 2005b; Poncet, 2006; Bao, et al. 2011, Shen, 2012 and 2013);
Temporal scale	1990 and 2000 (Beine et al., 2007); 1991-2000 (Lewer and Van den Berg, 2008); 1980-2004 (Hooghe et al., 2008); 1990-2000 (Pedersen et al., 2008); 1995-2004 (Warin and Svaton, 2008); 1986-2006 (Marques, 2010)	1985-90 (Poncet, 2006; Bao, et al. 2011), 1990-95 (Poncet, 2006); 1995-2000 (Cai and Wang, 2003; Fan, 2005b; Bao, et al. 2011); 2000-2005 (Bao, et al. 2011); 1985-2000 (Shen, 2012) and 1985-2005 (Shen, 2013).
Data sources	OECD, World Bank, UN, ILO and IMF publications (Pedersen et al., 2008); Eurostat NewCronos Database (Marques, 2010); OECD (Warin and Svaton, 2008; Hooghe et al., 2008; Lewer and Van den Berg, 2008); Census and register data from many OECD countries (Beine et al., 2007)	1990 and 2000 censuses (Fan, 2005b; Shen, 2012 and 2013); 1990 census and 1995 1% sample (Poncet, 2006), 1990 and 2000 censuses and 2005 1% sample (Bao, et al. 2011); 2000 census (Cai and Wang, 2003)
Model methods	OLS (Beine et al., 2007); Scaled OLS (Lewer and Van den Berg, 2008); general linear mixed modelling (Hooghe et al., 2008); Weighted least square and generalised estimating	OLS (Cai and Wang, 2003; Fan, 2005b; Bao, et al. 2011), OLS with fixed effects (Poncet, 2006); multi-level Poisson (Shen, 2012)

	equations (Pedersen et al., 2008); Kmenta-Parks method (Warin and Svaton, 2008); OLS with panel corrected standard errors (Marques, 2010)	
Dependent variable	Flow (Lewer and Van den Berg, 2008); inflow (Beine et al., 2007; Hooghe et al., 2008; Marques, 2010); out-migration rate (Pedersen et al. 2008); Inflow-rate (Warin and Svaton, 2008),	Flow (Fan, 2005b; Shen, 2012), destination choice (Poncet, 2006), out-migration rate (Cai and Wang, 2003; Bao, et al., 2011)
Population size	+ (Hooghe et al., 2008; Pedersen et al. 2008)	+ (Fan, 2005b; Shen, 2012)
Distance	- (Lewer and Van den Berg, 2008; Pedersen et al., 2008; Warin and Svaton, 2008; Marques, 2010)	- (Cai and Wang, 2003; Fan, 2005b; Poncet, 2006; Bao, et al. 2011; Shen, 2012)
Contiguity	+ (Pedersen et al., 2008); - (Warin and Svaton, 2008)	+(Poncet, 2006)
GDP per capita	- (Pedersen et al., 2008; Marques; 2010); + (Beine et al., 2007;Warin and Svaton, 2008; Marques; 2010)	+ or - (Fan, 2005b)
Income (or wage) difference	+(Marques; 2010)	+(Poncet, 2006; Bao, et al. 2011);
Unemployment rate	- (Hooghe et al., 2008, Pedersen et al., 2008; Warin and Svaton, 2008; Marques, 2010)	+ (Bao, et al. 2011);- (Poncet, 2006; Cai and Wang, 2003);
Stock of migrants (network effect)	+ (Lewer and Van den Berg, 2008; Pedersen et al., 2008; Marques, 2010);	+ (Cai and Wang, 2003; Fan, 2005b; Bao, et al. 2011)
Migration restriction	+ or - (Marques, 2010)	+(Bao, et al. 2011)

Notes: + (positive), – (negative)

Table 3. Macro-variables used for models

Type	**Variable**	**Description**	**Sources**
Dependent variable	M*ij*	Flow of gross migrants from origin *i* to destination *j*	The MIMOSA project (the EU) 2007, Chinese 2005 1% inter-census [b].
Geographical variables	D*ij* C*ij*	Straightline distance between the capitals of countries (EU) and railway distance between the capitals of provinces (China) in kilometres Spatial relationship between origin site *i* and destination site *j*: 1 (when adjacent) and 0 (not adjacent)	Google Online Geocoder, China national train time tables Derived from maps
Political variables	T*j* or Hk*j*	Tj - The immigration policy at destination *j*, measuring the effect of political regulations to migrant flow, transitional provisions in the EU case; or Hkj in the case of China as ratio of migrants with permanent local Hukou to total migrant population at the receiving province.	European Commission website[a]; Chinese population census of 2000[b].
Economic variables	G*ji* E*ji* W*ji*	Ratio of GDP per capita between destination and origin units; Ratio of employment rate between destination and origin units; Ratio of average wages between destination and origin units	Eurostat website[c]; National Bureau of Statistics of China website[b]
Demographic variables	P*i* or P*j*	Total population at origin or destination units	Eurostat[c]; National Bureau of Statistics of China website[b].
Network variables	N*ij*	Flow of migrants (for the previous 5 years) from the origin to the destination unit, for example flow of migrants between members of the EU prior to 2002 and between provinces in China prior to 2000.	Eurostat website[c]; National Bureau of Statistics of China website[b].

Notes: *a* (http://ec.europa.eu/); *b* (http://www.stats.gov.cn/);
c (http://epp.eurostat.ec.europa.eu/)

Table 4. Statistics of migration flows (M_{ij}) in the EU and China

Statistics	EU (2007)	China (2005)
Num. of observations	702 (27*26)	930 (31*30)
Min	0 (15 observations)	0 (96 observations)
Max	124,171	3,519,121
Mean	2,524	54,432
Standard deviation	8,861	220,457
Variance	7.86e+07	4.87e+10

Table 5. Measures of gross migrants flow

Index	China	EU
Average distance (D') (in *km*)	1317	679
Distance decay from negative binomial Regression *b* (co-efficient and alpha)	-1.231876 (2.681344)	-.875811 (1.69)
Connectivity (MC)	0.897	0.979
Inequality (IN)	0.711	0.673
Co-efficient of variation (CV)	4.05	3.51
Index of dispersion (DI)	0.641	0.628
Migration effectiveness (EI*)*	72.2	22.7

Table 6. Summary statistics of all variables created

Variables	Min	Max	Mean	St. Dev.
EU (702)				
Pi	405,006	82,437,995	18,267,664	22,674,119
Di,j	34 (km)	2,342 (km)	887 (km)	463
Gj,i	0.12	8.21	1.234	0.9
Wj,i	0.04	22.8	2.34	3.4
Ej,i	0.72	1.39	1.007	0.12
Ni, j	0	615,382	10,597	44,132
Tj	1	8	3.6	2.18
China (930)				
Pi	39,873,708	43,970,633	41,816,824	893,071
Di,j	143 (km)	5884 (km)	2035 (km)	1067
Gj,i	0.08	13.07	1.38	1.31
Wj,i	0.38	2.6	1.08	0.45
Ej,i	0.2	5	1.07	0.456

N i,j	0	2,495,498	41,753	168,603
Hj	0.05	0.54	0.33	0.12

Table 7. Gravity models calibrated by negative binomial regression

Variable	**EU**	**China**
M_{ij} – flow		
Independent var.	Co-efficient (t-value)	Co-efficient (t-value)
LnP_i	.7129059 (21.78) ***	.7321094 (9.72) ***
LnP_j	.6799997 (21.66) ***	-.036922 (-0.58)
$LnD_{i,j}$	-.7617719 (-11.29) ***	-1.320609 (-11.67) ***
Constant	-10.41525 (-12.23) ***	8.308329 (4.27) ***
alpha	1.670384	3.338182
Log likelihood	-5346.9853	-9885.0549
Pseudo R^2	0.0585	0.0115
Obser. Num.	702	930

*Note: *** at 1% significance (confidence) level*

Table 8 Full models of the EU and China calibrated by fixed-effect negative binomial regression

Variable	EU (Fixed-Effect)	CHINA (Fixed-Effect)
Mij – dependent var.		
Independent vari.	Co-efficient (robust std error)	Co-efficient (robust std error)
Cij	.3812 (.17) **	1.195(.152) ***
Ln Di,j	-.231 (.102) **	-.498(.081) ***
Ln Gj,i	-1.108 (.167) ***	.75 (.06) ***
Ln Ni,j	.429 (.021) ***	.369 (.037) ***
Ln Ej,i	2.027 (.712) ***	.206(.096) **
Ln Tij	-.057(.0645)	.376 (.042) ***
Wald chi2	31001.73	111721
Prob > chi2	0.000	0.000
Log likelihood	-5169	-9291
Obs. Num.	702	930

*Note: *** 1% significance, ** 5% significance, * 10% significance.*

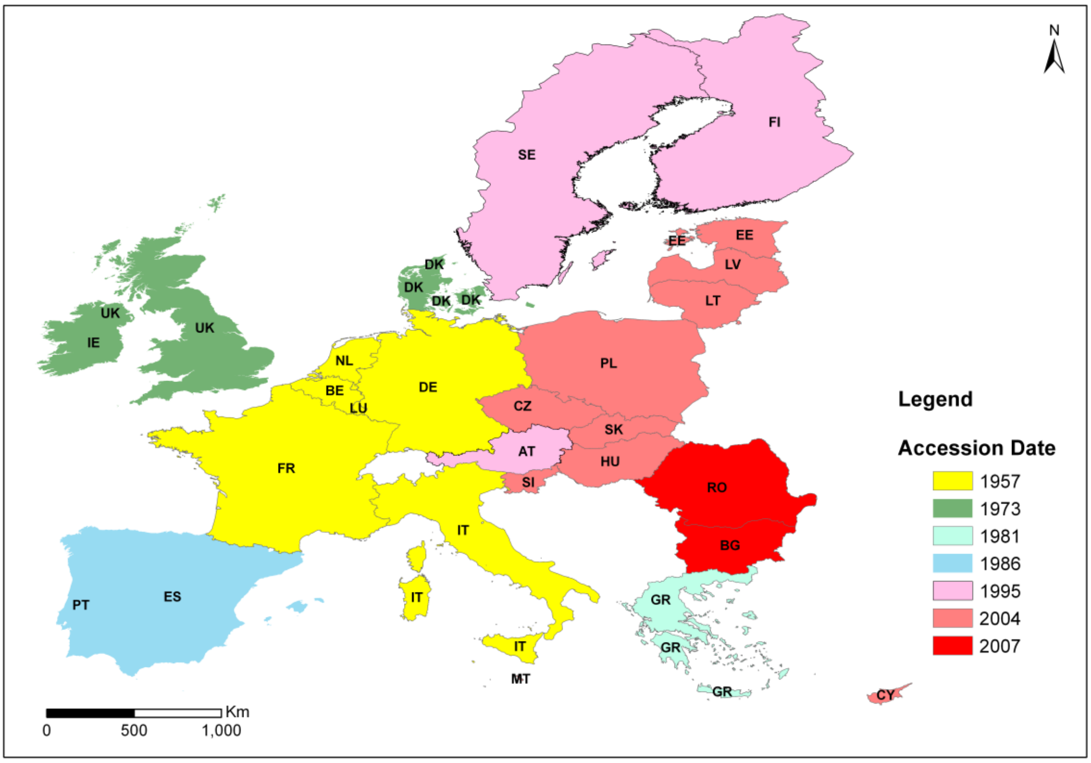

Figure 1. Members of the EU and its enlargement process

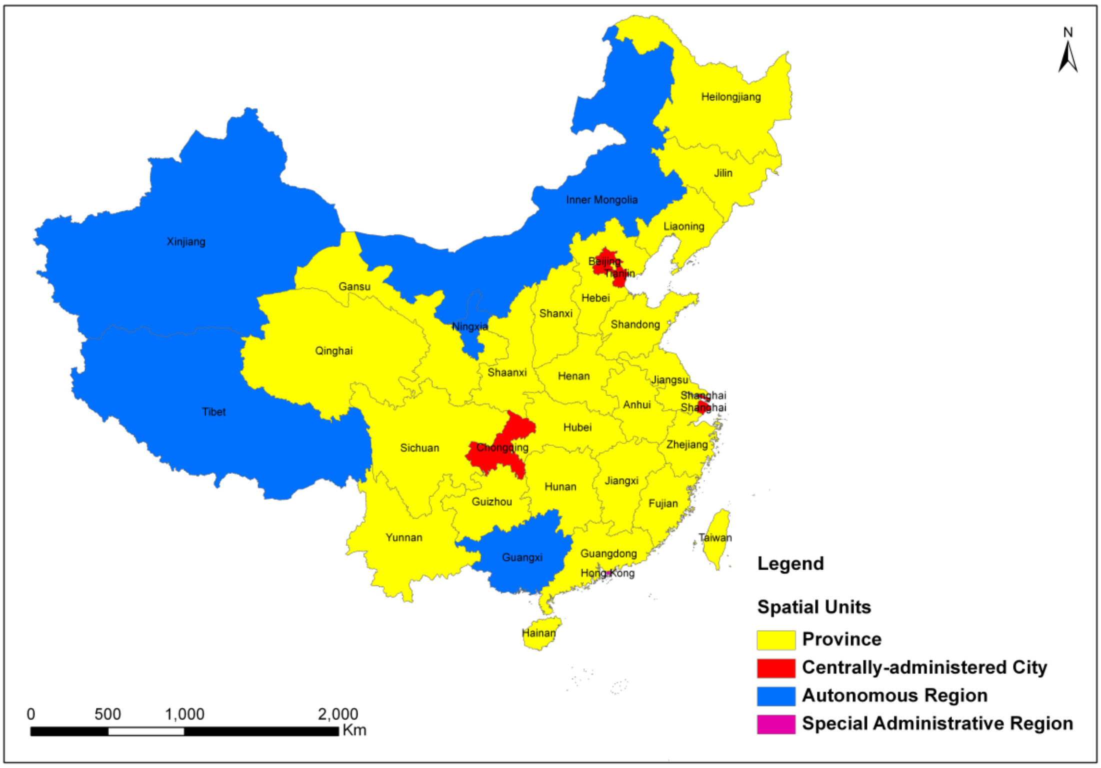

Figure 2. Administrative units (provinces level) in China

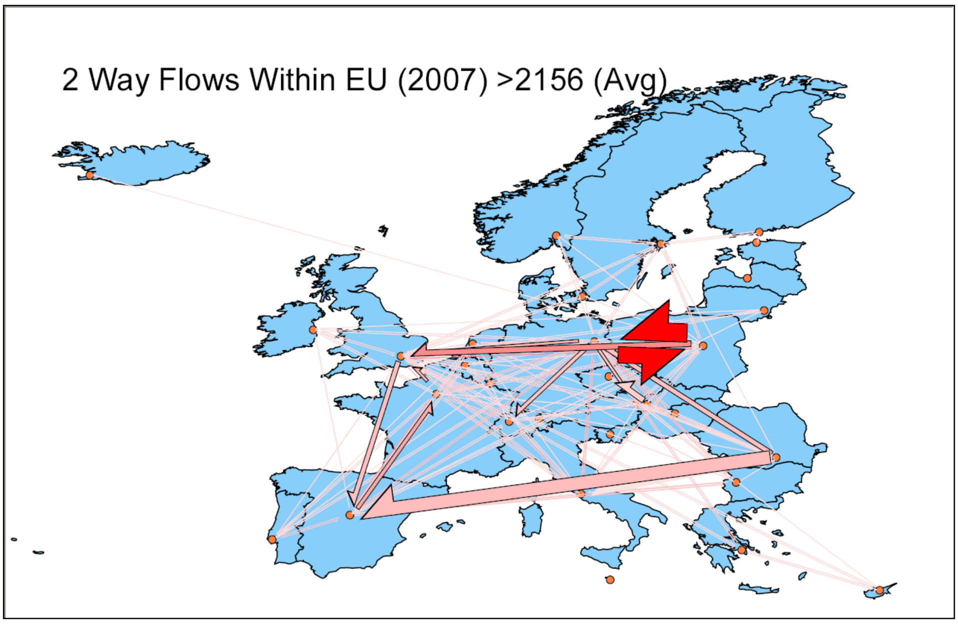

Figure 3. Flows of gross inter-state migrants across the EU27 in 2007

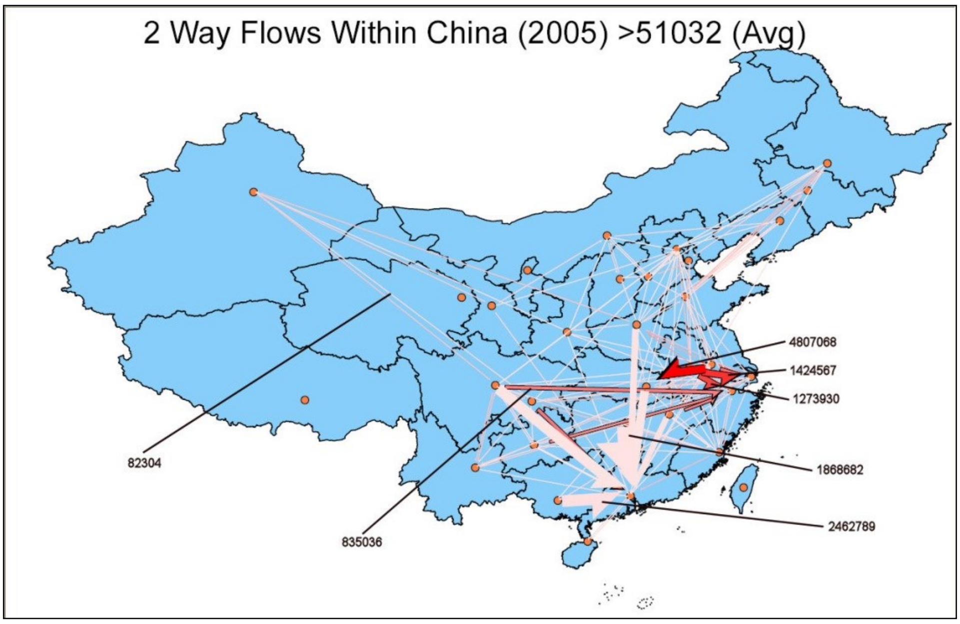

Figure 4. Flows of gross inter-provincial migrants across China in 2005

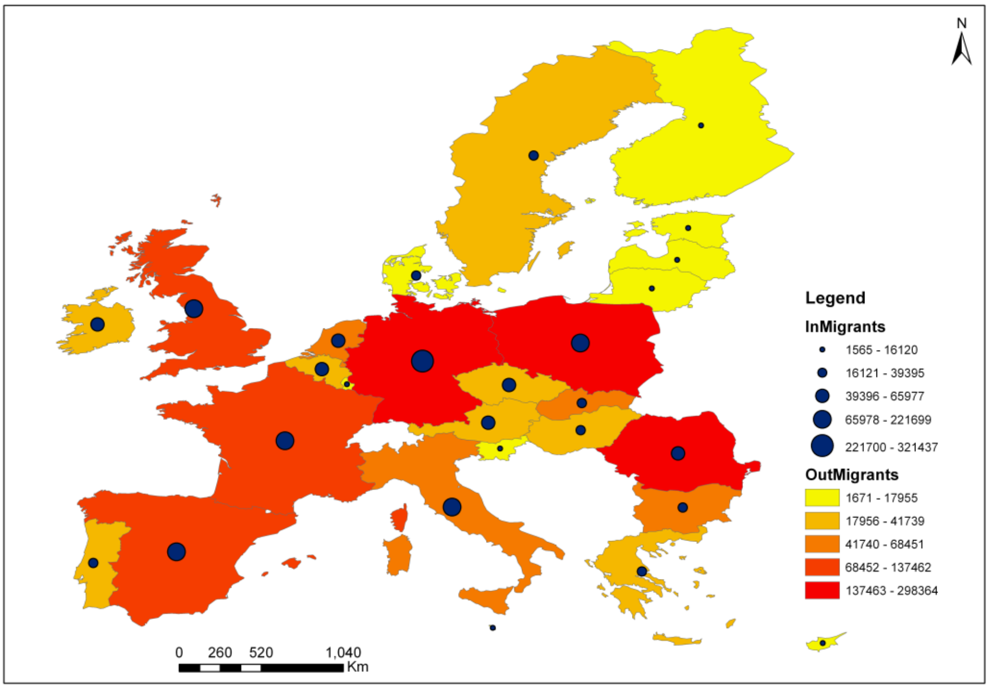

Figure 5. In-migration and out-migration across the EU in 2007

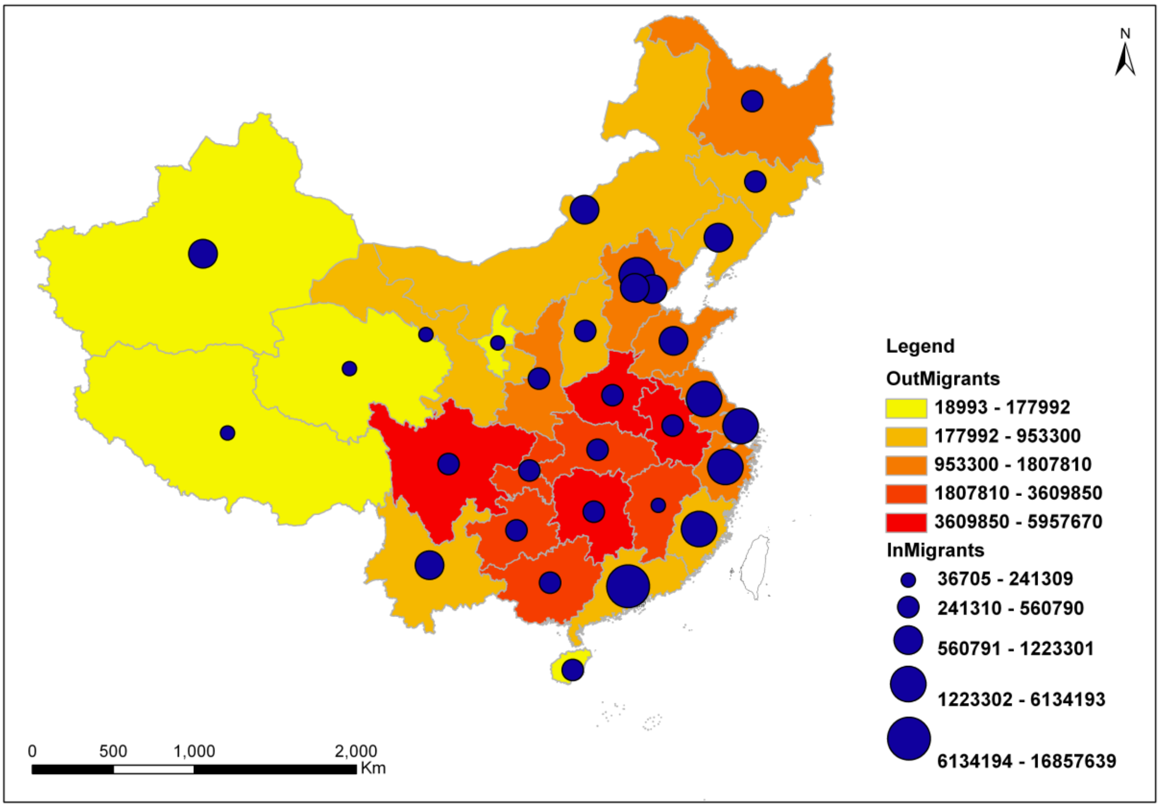

Figure 6. In-migration and out-migration across China in 2005

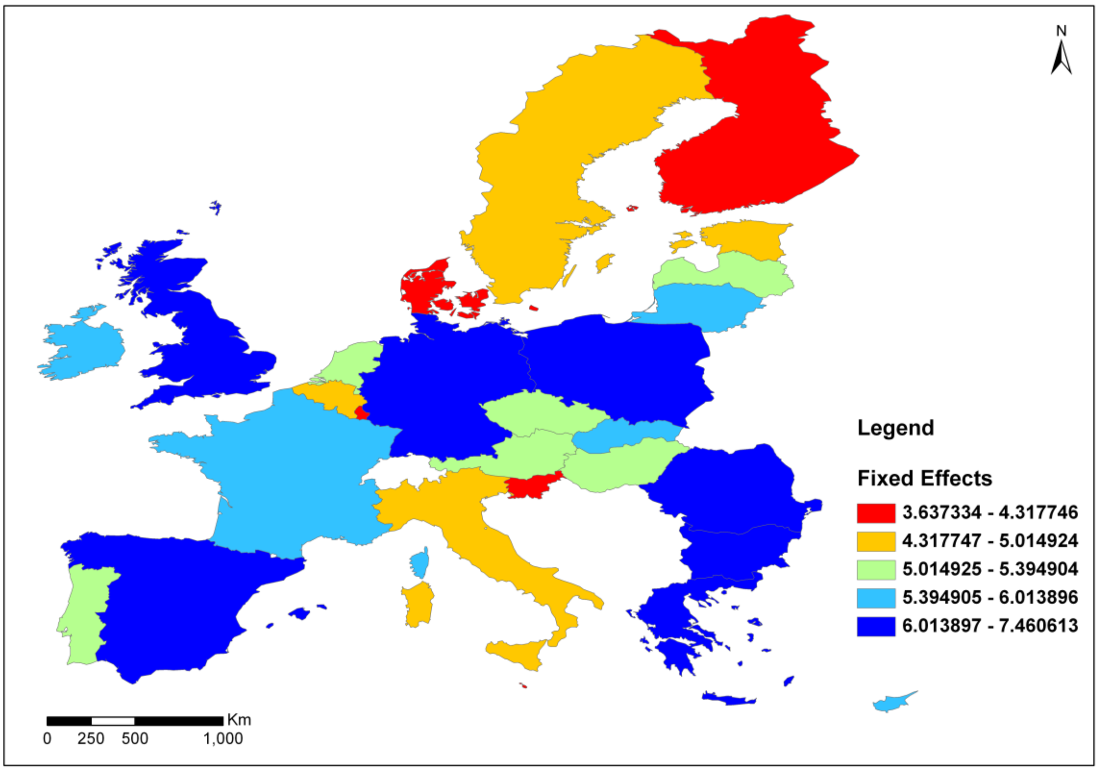

Figure 7. Fixed-effects of the EU model

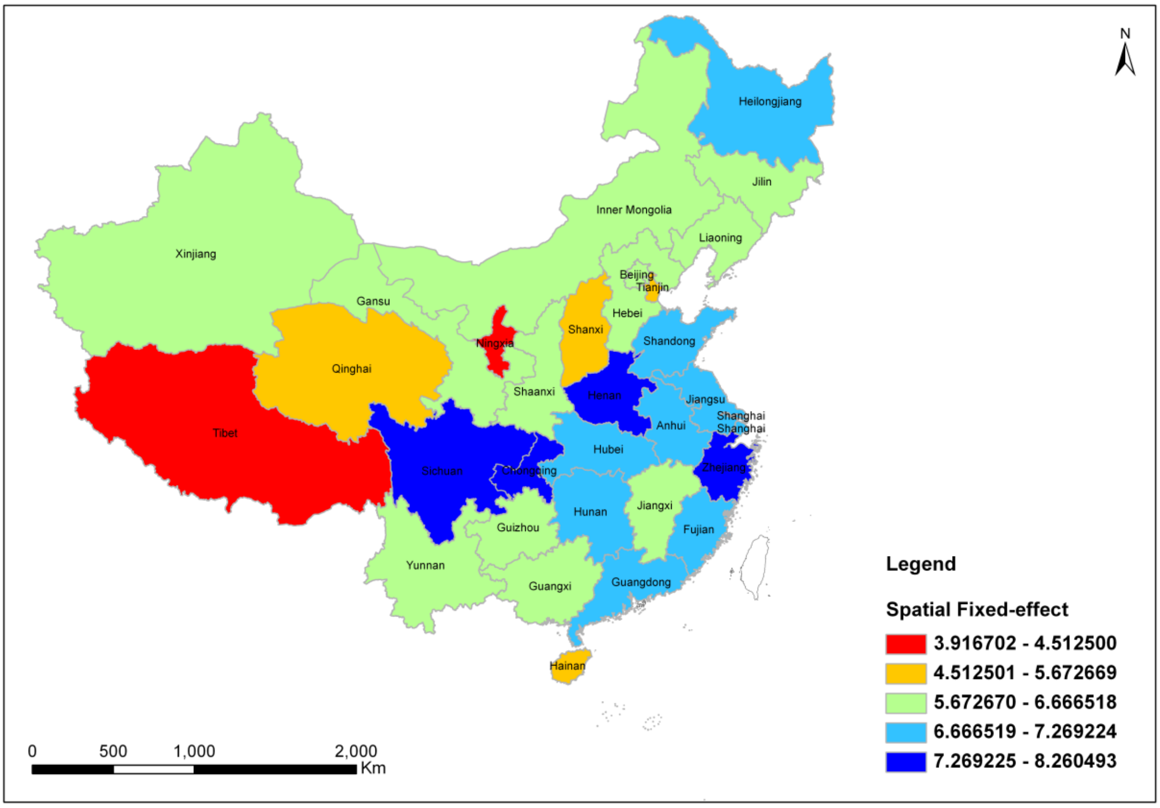

Figure 8. Fixed effects of the China model

Appendix A: Maps of the variables defined

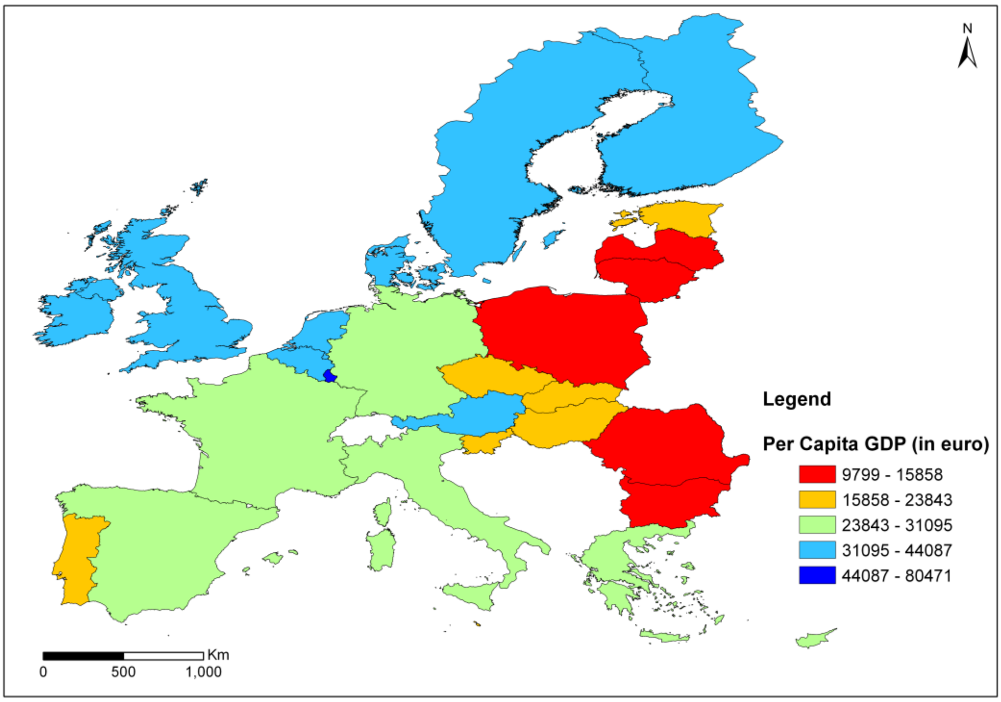

Figure A1. Per capita GDP distribution in the EU for 2006

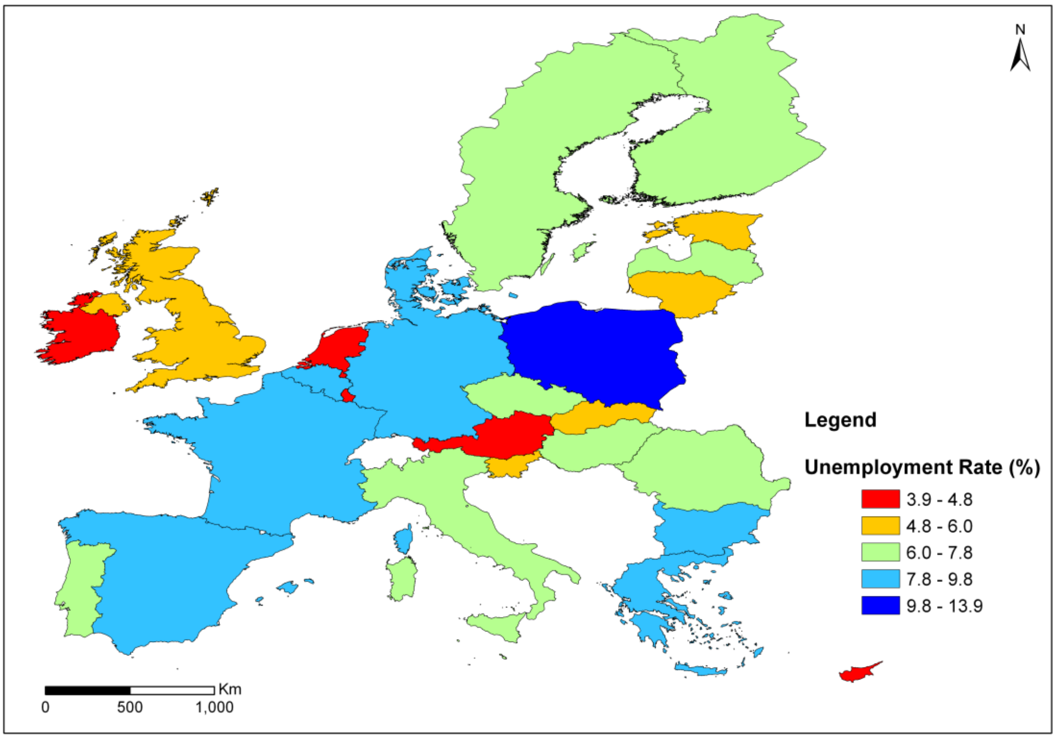

Figure A2. Unemployment rate of member states across the EU in 2006

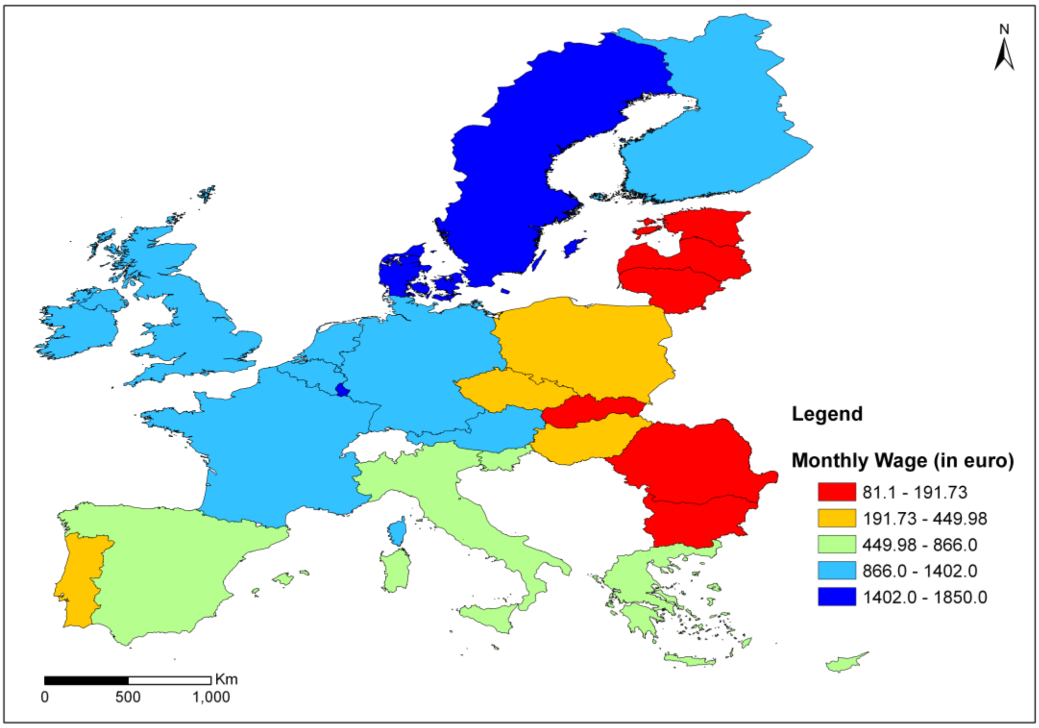

Figure A3. Average wage distribution in the EU for 2006

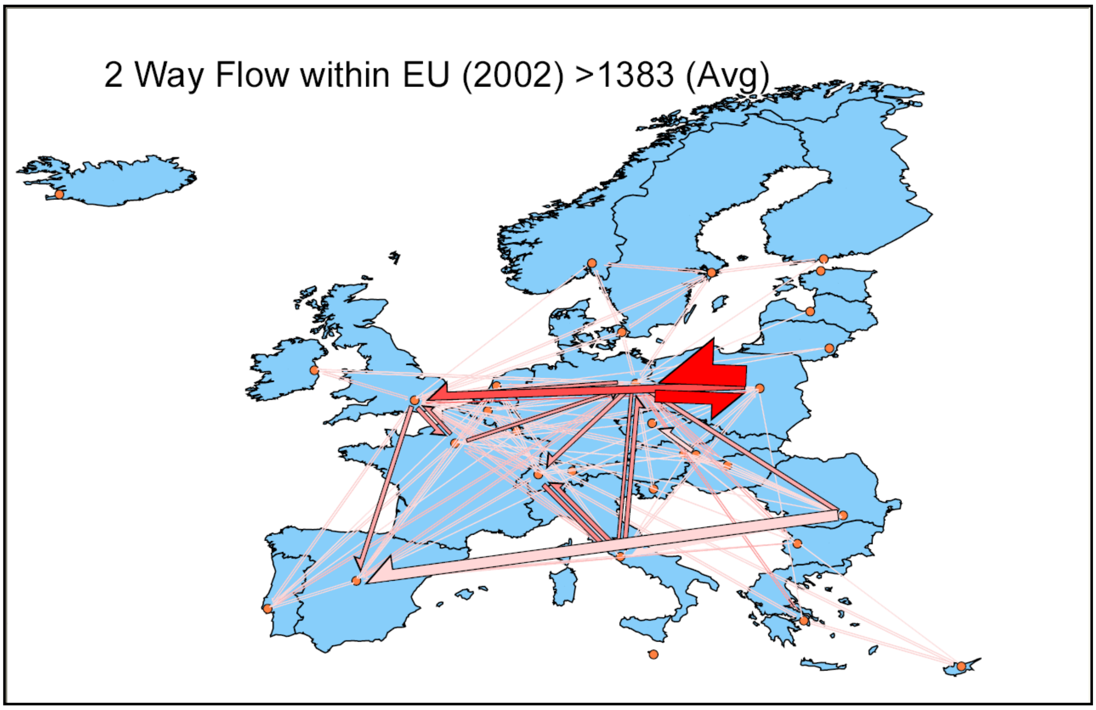

Figure A4. Network effect (flow of gross migrants in 2002) in the EU

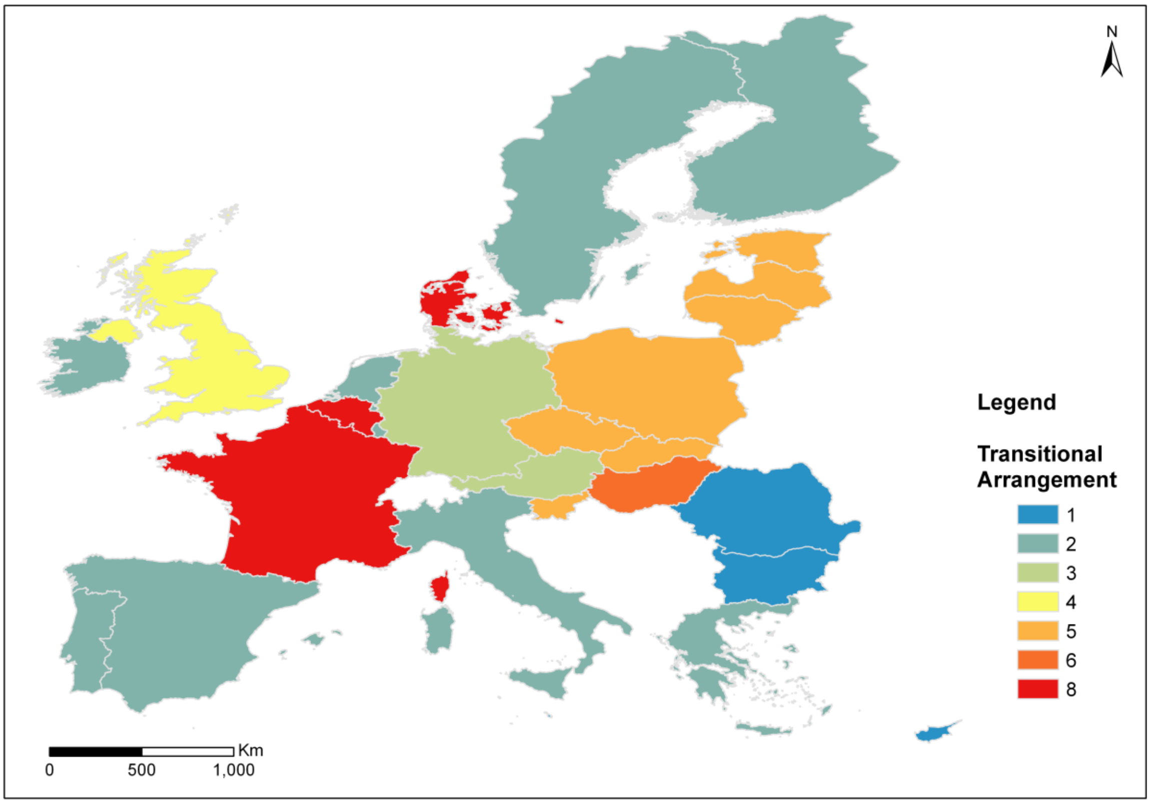

Figure A5. Restriction ranking of transitional provisions in the EU for 2006

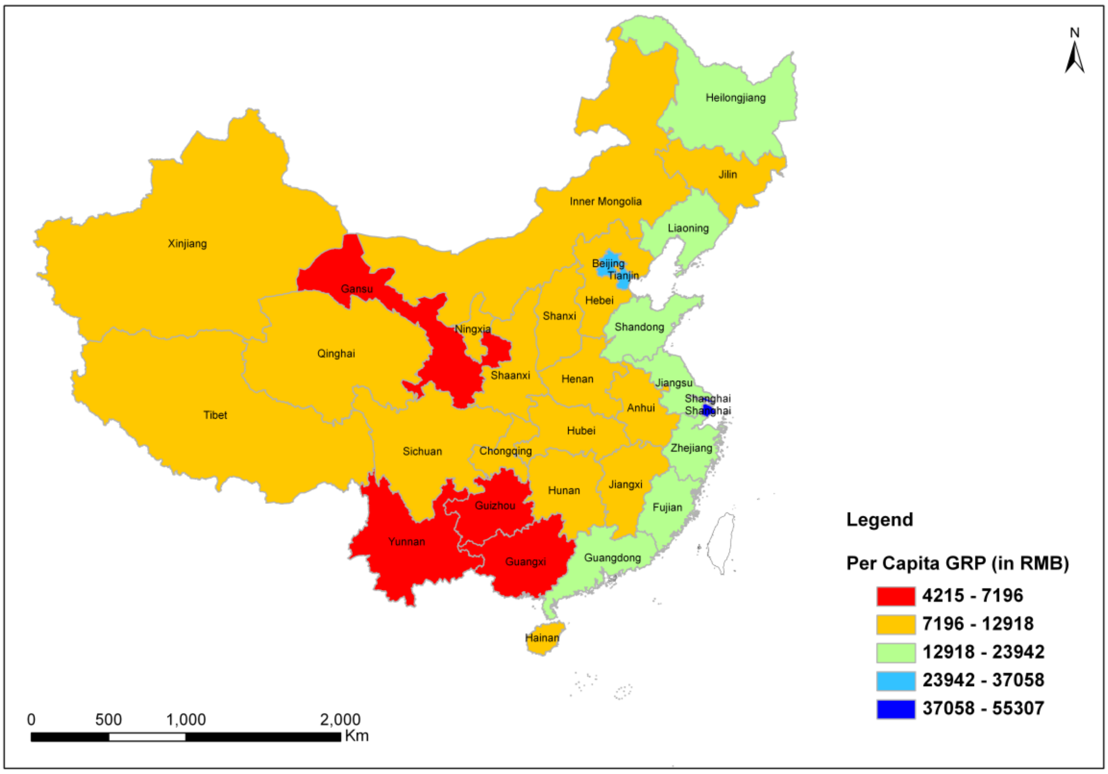

Figure A6. Per capita GRP distribution in China for 2004

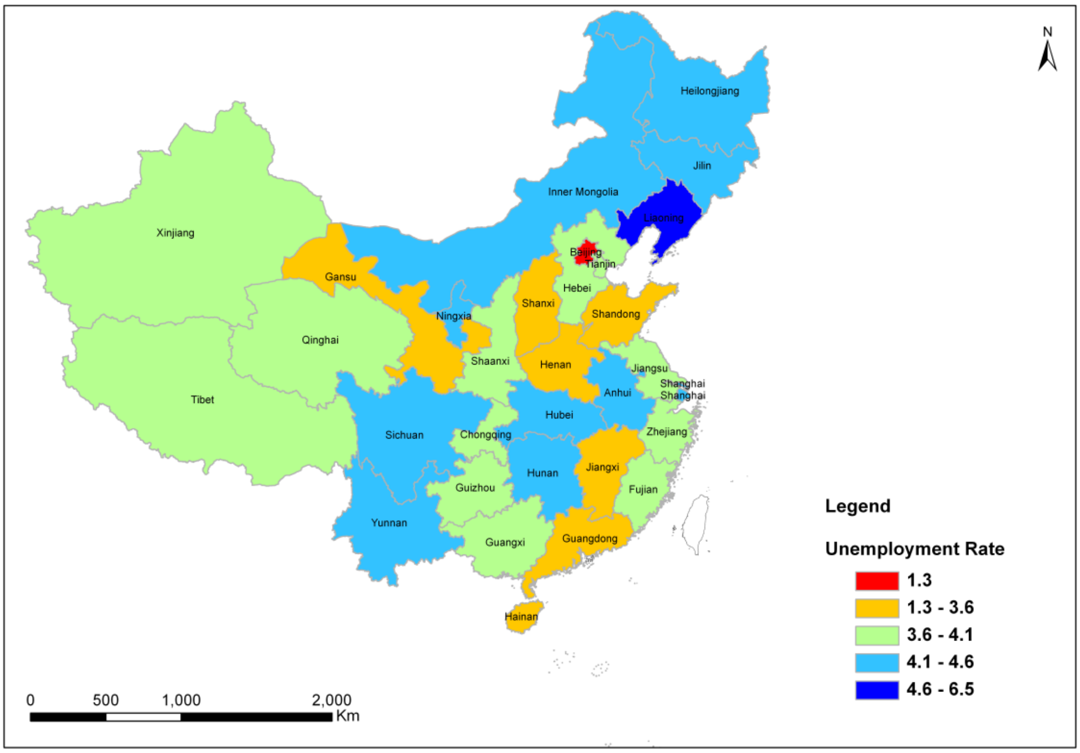

Figure A7. Percent unemployment rate in China for 2004

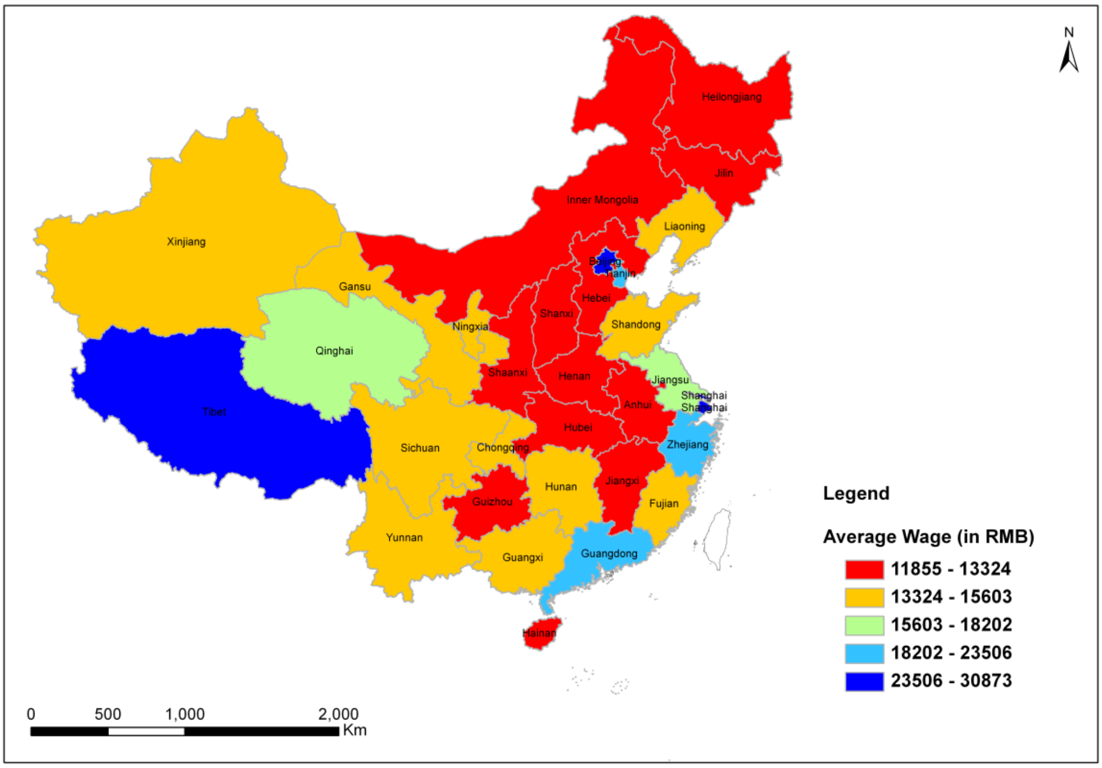

Figure A8. Average wage distribution in China for 2004

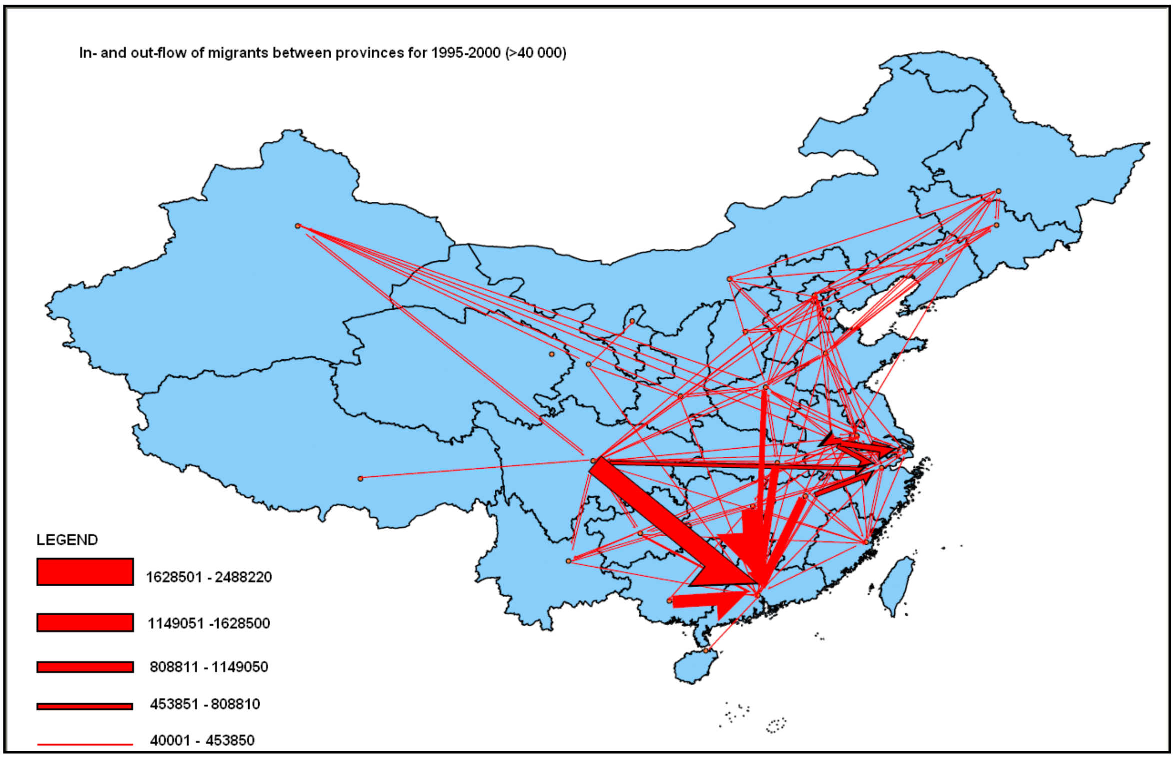

Figure A9. Network effect (flow of gross migrants in 2000) in China

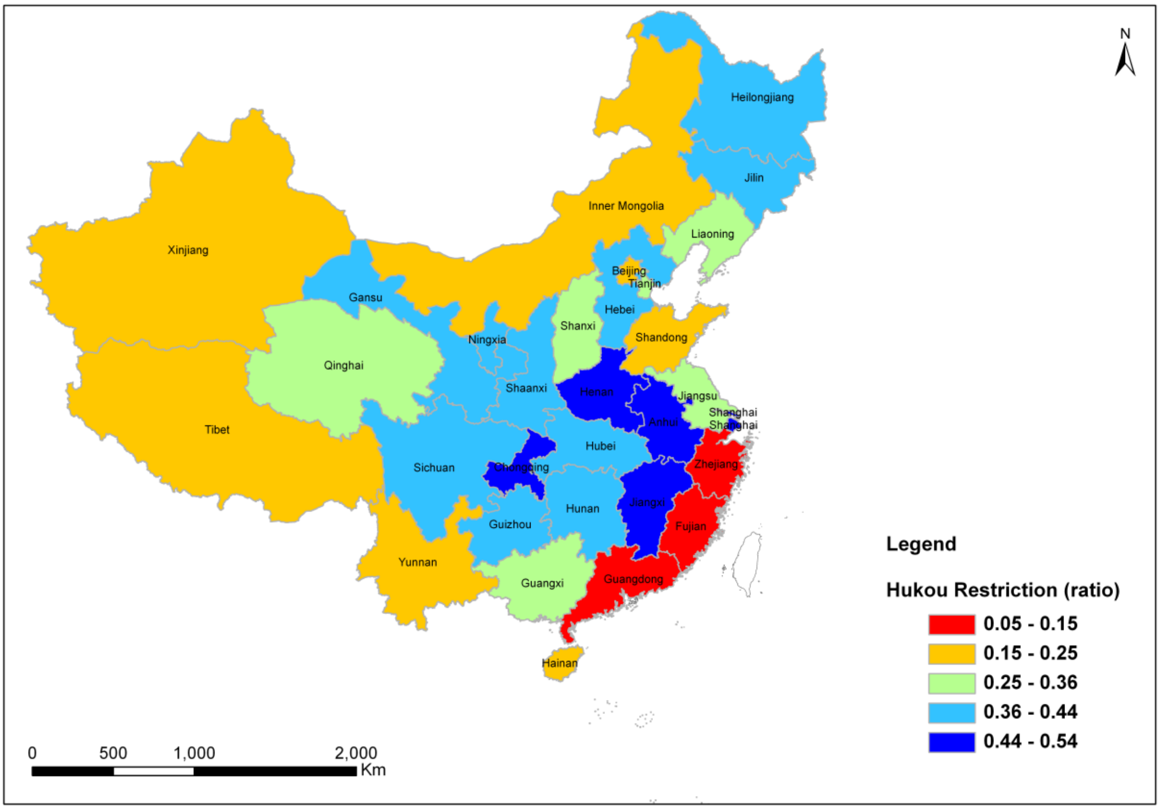

Figure A10. Ratio of migrants with permanent local Hukou to total migrant population.