

ORIGINAL ARTICLE

Open Access

Where is the grass greener? A micro-founded model of migration with application to Guangdong

Peter Simmons* and Yuanyuan Xie*

*Correspondence:
peter.simmons@york.ac.uk;
yx518@york.ac.uk
Department of Economics,
University of York, York, UK

Abstract

This paper first develops a structural micro-founded model of aggregate net migration flow using matching ideas to study how migrants choose between multiple locations using multiple criteria. Migration should reduce inequality in the criteria. Most migration models either do not handle multiple criteria and locations or lack micro foundation. The model predicts that migration flows will be out of all but the top two ranked regions. The empirical work, which uses 1990–1999 Guangdong annual data, confirms this proposition and finds a high degree of common marginal effects of the criteria among 18 locations but also finds increasing regional inequalities.

JEL: F22 J61 O15

Keywords: Multiple net migration flows, Micro-founded migration model, Regional inequalities

1 Introduction

Economic models of migration can be put in a general framework in which each possible location is perceived by an individual to have costs and benefits (Krugman 1992). Migration of an individual occurs when the net gain from changing location exceeds the migration cost. Previous research has identified some key pull factors motivating migration as an expected income gain (Harris and Todaro 1970; Johnson 2003; Todaro 1969), improved employment opportunities (Fan 1996; Greenwood et al. 1986; Zhao 1997), benefits from the infrastructure of an alternative location both economic and cultural infrastructure/amenities (Chen et al. 2008; Docquier and Rapoport 2005), and marriage opportunities (Fan 2002; Fan and Huang 1998; Frutado and Theodoropoulos 2008; Seeborg et al. 2000). An individual may of course be motivated by more than one of these push-pull factors and different possible locations may provide different mixes of these factors. The economic costs to migration include the immediate transport cost but also other fixed and variable costs in the destination location like education for children (Meyerhoefer and Chen 2011; Plantinga et al. 2012), housing market conditions (Vermeulen and Ommeren 2009), regulatory and institutional barriers reflected in national immigration controls and systems of internal administrative control on migration like the hukou in China (Chan and Zhang 1999; Renard et al. 2011; Whalley and

Zhang 2007). Typically the costs will vary with destination but, (apart from the immediate transport cost) not with the original location. The combination of all these forces applied to each individual will determine who moves where. In a matching equilibrium, the population distribution between locations and the specific features of different locations are such that no individual has an incentive to move. Matching theory results by Ekeland et al. (2004) and Heckman et al (2002) in particular are relevant to our purpose in which heterogeneous individuals are sorted into different outcomes by a market system which allows for pricing by type of individual. Our approach applies some of the insights of this literature to the migration process, recognising that migration is essentially a disequilibrium event.

The aim of this paper is to develop a theoretical framework¹ which allows for a menu of alternative migration pull factors in a geographical domain with multiple possible destinations. Each individual currently resides in a particular location and has to weigh up the alternative menu of factors available in each different location, together with the costs of moving to that location. The outcome is a migration decision for each individual and, aggregating these over individuals in each current location, we derive the net migration flows between locations. Apart from the theoretical insights gained into modelling multiple motives for migration and multiple locations, our second main aim is to use this framework to generate an empirically applicable migration equation system. We then apply this to econometric analysis of intra-province migration patterns in Guangdong, one of the provinces of China with the highest such migration rate.

Our main theoretical result is that the interplay of multiple motives and multiple possible destinations for migration will lead to agglomeration. Individuals will typically want to move to a location seen as best for the combinations of its advantages. However individuals already living in such a favoured location, but whose individual circumstances have turned out poorly there, may wish to leave and move to the location offering the second best advantages. Our approach establishes these tendencies in quite a general framework. We add some assumptions to the framework which results in a net migration equation which can be empirically applied in a setting with multiple migration motives and multiple locations. We apply this to intra-province migration in Guangdong for the period 1990–1999 with 18 locations and 4 migration motives. We find that the approach represents the data quite closely and that there is a high degree of preference homogeneity between individuals in different locations. This implies that if moving costs are low, the equilibrium spatial population distribution should yield equality between locations in the factors which cause migration. However this is not the case for all factors, in fact regional inequality has been growing for these despite the high net migration flows. We conclude that spatial equilibrium in Guangdong has not yet been reached.

Section 2 briefly reviews existing knowledge of factors causing migration and empirical work on Guangdong, Section 3 develops the theoretical framework, Section 4 describes the data, outlines the econometric strategy and presents the empirical results. Section 5 contains a conclusion.

2 Literature review

Some of the comprehensive surveys of migration research recognize the multiple push-pull factors which drive it and that migrants face a choice between

alternative destinations. Greenwood (1997) studies the determinants of migration in developed countries, including characteristics both of places and of individuals and households. Taylor and Martin (2001) study the complexity of migration determinants and its impact in rural economies. In the context of climate driven migration, Lillioer and Broeck (2011) also recognize the variety of migration causes.

Widely noted pull factors that we subsequently use are:

- (i) **Expected Wage Income.**
In the classic Harris-Todaro model approach, migrants are motivated by the high wage and the high employment probabilities in the chosen destination. There are many subsequent applications of these two ideas, for example Johnson (2003) finds that rural-urban migrants in China move with an urban-rural wage gap of 50 – –70%. For employment probability, Greenwood et al. (1986) find an elasticity of about 0.5 of migrants to job vacancies, and also unemployment shocks often hit immigrant workers most heavily Brucker and Jahn (2008).
- (ii) **Expected Self-Employment Profits.**
In European countries, 10–25% of migrants establish themselves as self-employed in the destination (OECD 2010) and in China it rises to 40% or more (RUMICI 2007) or even 60% in Guangdong (Fan 1996, 1999, 2003). There is evidence that self-employed migrants are most influenced by the size of market and population of the destination since diverse social networks are important (Federician and Giannetti 2010; Kugler and Rapoport 2005). In the UK there is a high concentration of self-employed immigrants in London (Dustmann et al. 2007) and 47% of the national self-employed migrants live there (The Migration Observatory 2012).
- (iii) **Location Infrastructure.**
Available local public goods differ by location eg transport and communications, public health or education services, cultural aspects and these affect the perceived quality of life in locations (Rappoport 2008). Synergies and externalities have a similar effect (Chen et al 2008). Often these come just from the size of the population in the destination.
- (iv) **Female Migration & Marriage Motives.**
Globally around 50% of migrants are female (Piper 2005). Marriage is one important pull factor in the UK (accounting for 40% of migrant settlements during 2008–2010 (Charsley and Liversage 2011) and Asian (Fan 2002; Fan and Huang 1998; Zhu 2002. Another is employment opportunities which may be gender specific eg in textile industries (Seeborg et al. 2000). Marriage can also interact with employment prospects (Frutado and Theodoropoulos 2008).

An important issue is identifying and modelling the costs associated with migration between particular locations. Transport costs depend on the physical distance between locations (Poncet 2006). In several NELM models (Docquier and Rapoport 2005; Mesnard 2000), fixed costs of migration are used. Institutional and regulatory barriers such as visas or the hukou system in China Chan and Zhang (1999) create costs. Whalley and Zhang (2007) conclude that inter-province wage differentials caused by the hukou system impose a significant welfare loss. In Guangdong (our focus for empirical work) the majority of migrants initially hold an agricultural hukou (Fan 1996,

1999, 2003; Zhu et al. 2009) which means permanent migration will require a change of hukou.

3 Theoretical background

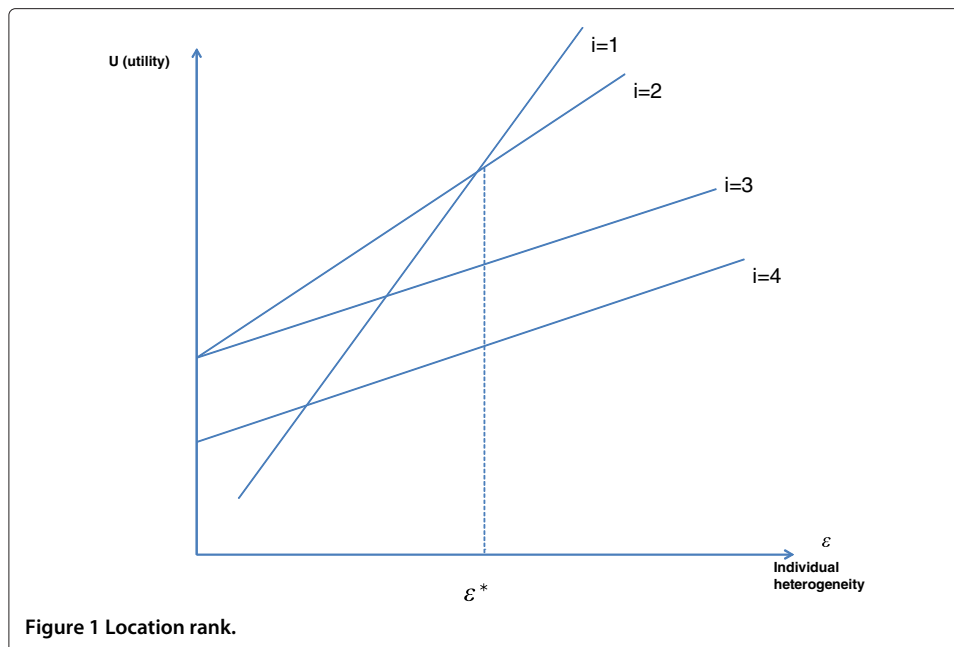
3.1 General framework

There are n individuals $h = 1..n$ who are each initially located in one of m different places $i = 1..m$. Each individual derives utility from K (continuously divisible) location specific factors² $f = (f_1, f_2, \dots, f_K)$. Individuals have utility $u(f)$ and by definition utility is increasing in each factor. Within each location there is a multivariate distribution of the factors across the individuals who reside there, so for example individuals h and l who both reside in the same location may respectively enjoy values of the factors $f = (f_{1h}, f_{2h}, \dots, f_{Kh})$ and $f = (f_{1l}, f_{2l}, \dots, f_{Kl})$. This generates intra-location differences. In addition locations differ in the distribution of factors so there are also inter-location differences, stressing the role of individual heterogeneity in line with Borjas (1999) and Ekeland et al (2004). For example the mean or the variance of factors realised may differ between locations (one location may be relatively rural with a low mean and variance of wages, another have a more mixed industrial structure with both a higher mean and a higher variance in wages). In terms of matching theory there is bunching in locations because two individuals who are both ex-ante best suited to a particular location experience different “luck” in accessing the attributes of the location eg by chance one individual may get a better job offer than the other even though they are of identical productivity. But on average one location may attract mainly low skilled workers whilst another attracts high skilled workers. Thus within location utility differences largely result from luck but between location differences from more deterministic heterogeneity in location and individual characteristics.

In a matching model, Ekeland-et al (2004) have derived a closed form solution for the equilibrium matching allocation when individual utility is linear in unobserved deterministic heterogeneity but quadratic in the location factors. Suppose that average utility derived from factors z is quadratic in z , $a'z + 1/2z'Bz$, and immigrants into the overall area have to choose their location. Individual deterministic heterogeneity, ε , causes differences in realised utility between individuals who choose the same location according to $\varepsilon a'z + 1/2z'Bz^3$. So the N locations can be ranked by each individual with the best location for ε being $\max_i \in N \varepsilon a'z^i + 1/2z^i Bz^i$ where z^i is the vector of factors in location i .

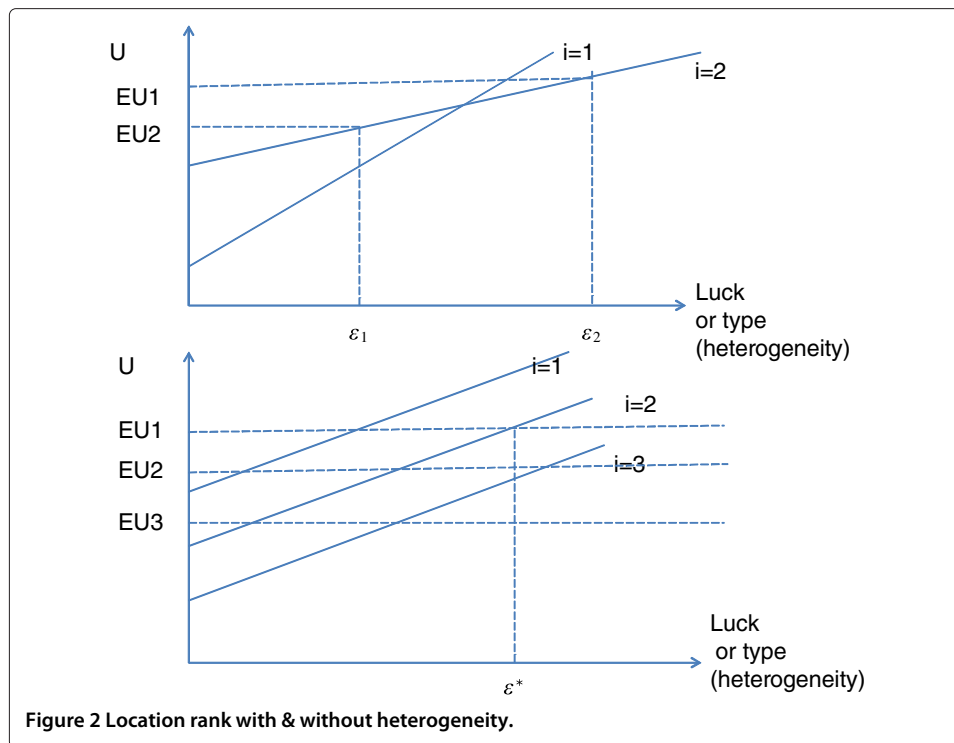
Figure 1 (Figure 1 location rank) shows the utility available to different individuals from locating in each of four different locations $i = 1, \dots, 4$. Locations 3 and 4 are dominated and optimal for no individual. Any individual with $\varepsilon > \varepsilon^*$ has location 1 as their top ranked destination with second best location 2. Any individual with $\varepsilon < \varepsilon^*$ has top ranked location 2 with second best choice either location 3 or location 1. Hence the population is sorted into one group who will migrate into location 1 and another who will migrate into location 2. If that is the whole story then each individual will locate in his best destination.

We have to adapt this to analyse movement between locations eg from one location to another within Guangdong. Each person within any given location compares their current circumstances with what they can expect to attain by a move to an alternative location, this interaction determines who moves where. An individual h currently in i could move to location j . If he moves, he does not know exactly what factor combination he will get in j because there is heterogeneity within a location in individual experience there. He has



to assess the range of payoffs he could get from different possible combinations x of the factors $f_{1x}^j, f_{2x}^j, \dots, f_{Kx}^j$ which might occur in j . We assume each individual assesses his gross benefit from moving to j as $Eu(f|j) = \sum_x \pi^j(f_{1x}^j, f_{2x}^j, \dots, f_{Kx}^j) u(f_{1x}^j, f_{2x}^j, \dots, f_{Kx}^j)$. Here $\pi(\cdot)$ has the interpretation that in j , h has the chance π of getting the factor combination $f_{1x}^j, f_{2x}^j, \dots, f_{Kx}^j$ in j . It is then natural for any individual to condition his probability distribution over factors in j on the mean level of the factors there which are observable to the individual, μ_1^j, \dots, μ_K^j , so that $\pi^j = \pi(f_{1x}^j, f_{2x}^j, \dots, f_{Kx}^j, \mu_1^j, \dots, \mu_K^j)$.

Again locations are ranked by each individual, but now the best off individuals in any location may prefer to stay where they are rather than to move, whilst those with the lowest standard of living in a location are the most likely to move. The heterogeneity within a location may partly depend on deterministic individual characteristics but also in large measure on differences in luck that individuals have experienced in the current location. We can model this by assuming that any individual who enters a new location has an equal chance of enjoying the mean standard of living there. They can also stay in their present location and enjoy their current standard. The first diagram of Figure 2 (Figure 2 location rank with & without heterogeneity) shows the utility distribution of different individuals in a given location as a function of their past experience (luck) and type on the horizontal axis, for example ranging from very unlucky to very lucky or from unskilled to skilled. The positively sloped lines show the current utility enjoyed from different types at present in that location. The horizontals E_1, E_2 show the mean utility level that any type of one location can expect from a move to the other location. Thus in location 1 all types with utility below currently below ε_1 will move to location 2 but all others will remain in location 1. Similarly all types in locations 2 with $\varepsilon < \varepsilon_2$ will move to location 1 but all others will remain. In fact in our data observed migrants actually have very similar individual characteristics in terms of age, education, marital status (RUMICI 2007) so presumably the within location heterogeneity is largely caused by past experience. Also in the data the observed city factors z in the data actually have strong dominance relations (the same city



empirically tends to come top on each factor in our sample period). This gives us the second diagram of Figure 2 (Figure 2 location rank with & without heterogeneity) in which the utility distributions of different locations never intersect and the mean utility of each location is unambiguously ranked. In this case below the best location, the worse off in each location all move into location 1, but the worst off in location 1 move to the second best location 2.

The implication is that with homogenous expected utility (preferences $u()$ and probabilities π_x^j) between individuals, all individuals in all locations will agree on j^* as being the location offering the highest $Eu(f|j)$. Any two individuals who presently have the same factors will agree on the new particular j^* location that offers the best standard of living $Eu(f|j^*)$ irrespective of where they are currently located (so long as they are not located in j^*).

However measured in commensurate utility terms, if the individual does move from i to j he also bears costs v_{ij} so the net expected benefit he could secure from a move to j is

$$NB_{jh} = \sum_x \pi_x^j u(f_{1x}^j, f_{2x}^j, \dots, f_{Kx}^j) - v_{ij} - u(f_{1h}, f_{2h}, \dots, f_{Kh})$$

The best alternative location for h is then $NB_{j^*h|i} = \max_j \{NB_{jh}\} | j = 1 \dots n | i$ and h will move to j^* if $NB_{j^*h|i} > 0$.

It is also natural to assume that the moving cost v_{ij} is additive in ij , that is $v_{ij} = c_j + c_i$. Costs based on distance between locations will satisfy this and so will any other fixed moving costs. There is little reason to suppose that leaving costs from exiting i depend on the destination so $v_{ij} - v_{ik} = v_{lj} - v_{lk}$. With an appropriate choice of units, this implies that v_{ij} is additive. Indeed the exit costs from a location may be very close to zero. With additivity of costs, the best destination $j^* = \max_j \{ \sum_x \pi_x^j u(f_{1x}^j, f_{2x}^j, \dots, f_{Kx}^j) - c_j \}$. If v_{ij} is small

too, then relative to the utility differential $Eu(f|j^*) - u(f_{1h}, f_{2h}, \dots, f_{Kh})$ will determine the best destination for individual h .

If an individual is currently in j^* any uncertainty about the combination of factors he faces there has already been resolved. However such an individual can still think about moving elsewhere and if he is currently low down in the distribution of attainable utilities in j^* , then he may have the prospect of a higher expected utility by moving to an alternative location. This location j^{**} will be the one which is judged second best by all individuals. That is j^{**} maximizes $\sum_x \pi_x^j u(f_{1x}^j, f_{2x}^j, \dots, f_{Kx}^j) - c_j$ over all j other than j^* .

How does this general framework lead to migration flows between locations? In any location i there is a distribution of the factors f defined by the cdf $G_i(f)$ with associated density $g_i(f)$. Define the lower sets of $u(\cdot)$ by $L(u^*) = \{f|u(f) \leq u^*\}$. Then there is a corresponding distribution of utilities in the location of $H_i(u^*) = \Pr(u(f) \leq u^*) = \int_{f \in L(u^*)} g_i(f) df$. All individuals in location i with $\sum_x \pi_x (\mu_1^j, \dots, \mu_K^j) u_x - c_j > u(f_1^i, \dots, f_K^i) + c_i$ will desire to move. If the exit cost c_i is zero, a proportion $H_i(\sum_x \pi_x (\mu_1^j, \dots, \mu_K^j) u_x - c_j)$ could increase their utility by a move to location j . With N_i individuals currently in location i , the number who could benefit from migrating from i to j will be $N_i H_i(\sum_x \pi_x (\mu_1^j, \dots, \mu_K^j) u_x - c_j)$. However since all individuals not presently in j^* agree that j^* offers the highest gain from all possible moves, the migration flow into locations other than j^* will be zero from origins other than j^* . So we derive aggregate immigration into j^* from all other locations of $I_{j^*} = \sum_{i \neq j^*} N_i H_i(\sum_x \pi_x (\mu_1^{j^*}, \dots, \mu_K^{j^*}) u_x - c_{j^*})$ and zero immigration into any other location.

However some individuals at present in j^* will have had unfortunate experiences there and will be low down in the utility distribution in j^* . They could gain from a move to the second best location. The number of individuals in j^* who see they can secure the highest improvement in their standard of living by a move to j^{**} is $N_{j^*} H_{j^*}(\sum_x \pi_x (\mu_1^{j^{**}}, \dots, \mu_K^{j^{**}}) u_x - c_{j^{**}})$.

Proposition 1. *If individual preferences and chances of success in any destination location are common, and if the migration cost is additive with exit costs being very low relative to entry costs, migration flows will be out of all but two locations. The two locations with immigration will be ranked as the top two in terms of the overall standard of living.*

This is a general abstract formulation and specializations of it are of interest, either to derive clear cut theoretical properties of migration flows like comparative statics with respect to characteristics of locations, including inequality in the distribution of factors within locations, or for empirical application. One natural specialization that we will use for empirical work is to assume that $H_i(\cdot)$ is linear ($H_i(\sum_x \pi_x (\mu_1^j, \dots, \mu_K^j) u_x - c_j) = \sum_k (a_k + b_k \mu_k^j) - c_j$). It is obviously very convenient and can arise in several scenarios:

(i) Specialized preferences

Suppose that in any location the population partitions into groups who are each affected by a single different factor. For example one group is employees who are motivated solely by labour earnings, another group are disabled individuals who are heavily dependent on local public goods like health care. Each individual within the group for which the k th factor is the sole determinant of utility judges alternative locations solely in terms of that factor. All individuals in this group in

any location judge the particular location $j^*(k) = \arg \max_j \sum_x \pi_x(\mu_k^j) u(f_{kx})$ as the most desirable in which to live. There is a distribution of the k th factor within location i which generates a utility distribution for this group of $H_{ik}(u(f_k))$. From each location i a proportion of individuals $H_{ik}(\sum_x \pi_x(\mu_k^j) u(f_{kx}))$ in the k -group will want to move to $j^*(k)$.

Taking each group in turn, for each factor there may be a different top ranked location $j^*(k)$ and a particular location may be top ranked on more than one factor. The immigration into any location j will be the sum of those individuals k in other locations i who judge j as the best destination on any one of the K factors. So we get immigration into j of $\sum_{i \neq j} \sum_k N_{ik} H_{ik}(\sum_x \pi_x(\mu_k^j) u(f_{kx}) - c_j)$ where N_{ik} is the number of individuals living in i who are motivated only by the k th factor. If the distribution of the factor within groups is uniform and so the cdf is linear, this reduces to $\sum_{i \neq j} \sum_k N_{ik} (a_k + b_k \mu_k^j) - c_j$. In addition there will be the second choice individuals who are presently in the top location for a particular factor that determines their standard of living, but whose present state on that factor is so unfavorable that they would be better off moving to the second best location on that factor.

(ii) Complementary preferences:

If $u(f_{1h}, f_{2h}, \dots, f_{Kh}) = \min(f_{1h}, f_{2h}, \dots, f_{Kh})$ then effectively each individual is constrained by a single factor (ignoring ties) in their current position. The utility they anticipate from any alternative location is $\sum_x \pi_x^j \min(f_{1x}^j, f_{2x}^j, \dots, f_{Kx}^j)$. We can partition the support of the distribution of f into regions R_k in which the k th factor is critical (for simplicity in the argument ignoring ties, which have minimal probability) so that eg for any $(f_{1x}^j, f_{2x}^j, \dots, f_{Kx}^j) \in R_k, f_{kx}^j = \min(f_{1x}^j, f_{2x}^j, \dots, f_{Kx}^j)$. If Π_k^j is the probability of the region of the support in which the k th factor is the constraining factor on utility, we can write $\sum_x \pi_x^j \min(f_{1x}^j, f_{2x}^j, \dots, f_{Kx}^j) = \sum_k \Pi_k^j \sum_{k \in K} \pi_{k|k} f_k^j$ where $\pi_{k|k}$ is the probability of particular values of the k th factor given that this factor is the minimal constraining one. For example an individual is affected by all factors but needs both a high wage and good infrastructure in fixed ratios. If the ratio is not met then his standard of living is set by the lower of the two. Hence his expected utility is determined by the chance that each of the two factors is constraining, and then within the region where one factor is critical, the expected value of the utility of that critical factor.

Under the assumption that $H_i(\sum_x \pi_x(\mu_1^j, \dots, \mu_K^j) u_x - c_j) = \sum_k (a_k + b_k \mu_k^j) - c_j$, aggregate immigration into j^* from all other locations is

$$M_{j^*} = \sum_{i \neq j^*} N_i H_i(\sum_x \pi_x(\mu_1^{j^*}, \dots, \mu_K^{j^*}) u_x - c_{j^*}) = \sum_{i \neq j^*} N_i (\sum_k (a_k + b_k \mu_k^{j^*}) - c_{j^*})$$

However if j^{**} is the second best location ($j^{**} = \arg \max(\sum_x \pi_x(\mu_1^{j^*}, \dots, \mu_K^{j^*}) u_x - c_j | j \neq j^*)$) then

$$N_{j^*} H_{j^*}(\sum_x \pi_x(\mu_1^{j^{**}}, \dots, \mu_K^{j^{**}}) u_x - c_{j^{**}}) = N_{j^*} (\sum_k (a_k + b_k \mu_k^{j^{**}}) - c_{j^{**}})$$

will leave the top location j^* seeking better fortune in j^{**} . Hence we can write the full set of net migration flows between locations as

$$\begin{aligned}
 NM_{j^*} &= \sum_{i \neq j^*} N_i (\sum_k (a_k + b_k \mu_k^i) - c_j) - N_{j^*} (\sum_k (a_k + b_k \mu_k^{j^*}) - c_{j^*}) \\
 NM_{j^{**}} &= N_{j^*} (\sum_k (a_k + b_k \mu_k^{j^*}) - c_{j^*}) - N_{j^{**}} (\sum_k (a_k + b_k \mu_k^{j^{**}}) - c_{j^{**}}) \\
 NM_i &= -N_i (\sum_k (a_k + b_k \mu_k^i) - c_{j^*}) \text{ for } i \neq j^*, j^{**}
 \end{aligned}
 \tag{1}$$

4 An empirical application to Guangdong

China's rapid development has been largely regional with construction of infrastructure and establishment and growth of an industrial base concentrated in particular areas. Resulting regional inequalities have stimulated migration, although the hukou system has acted as a migration barrier of variable force. Guangdong is a Chinese province close to Macao and Hong Kong, which has attracted government financial incentives for development and high FDI from Hong Kong. Growth has concentrated around the Pearl River, triggering high levels of intra-province migration (2.7 times higher than its already high inter-province immigration). We use 18 Guangdong city areas as our location units⁴ for the period 1990–1999. The cities are heterogeneous: Guangzhou (1), Shenzhen (2), Zuhai (3), Dongguan (10), Zhongshan (11) and Foshan (13) are industrialised Pearl River cities. Dongguan (10), Zhongshan (11) and Foshan (13) are traditional industrial centres and Shenzhen (2) and Foshan (13) are also Special Economic Zones (Nishitaten 1983) with government incentives. The cities further from the Pearl River nexus, Shaoguan (5), Heyuan (6), Meizhou (7), Huizhou (8) and Shanwei (9) in the north, Zhanjian (15), MaoMing (16), Yangjiang (14) and Qiangyuan (18) in the south, show lower net migration and some have net emigration. Finally there are cities (22) and (23) formed by merging the administrative units of (4), (19), (20) and (17), (21) (Figure 3 Map of Guangdong). The merged city has very low net migration.



Figure 3 Map of Guangdong.

For each city and year the variables we measure are

- (i) permanent net migration generally involving a change in hukou (NM),
- (ii) the population (both genders and all hukou types) (P)
- (iii) urban employment (E) and urban wage (W_u) in the top three permanent sectors (state owned, urban collective owned and other units)
- (iv) rural income per capita (W_r)(defined as the ratio of gross agricultural output in rural primary industry to rural primary industry employment)
- (v) a city specific consumer price index Π used to derive the real urban and rural wages $w_u = W_u/\Pi, w_r = W_r/\Pi$
- (vi) capital stock K derived from an initial stock, foreign direct investment flows, and a city specific depreciation rate⁵.

Table 1 (Table 1 Average value of key variables) describes the different city characteristics (the details of those key variables are defined in Table 2 Description of key variables): the net migration population ratio (NM/P), w_u, w_r , urban hukou holders as a % of the city population (Urbanhukou/ P), the late marriage rate⁶, capital stock per city inhabitant (K/P), the number of single females, the population (both in millions of people) and the city size in million square metres. The first three cities are the key Pearl delta cities and are the most urbanized, physically small, highly industrialized and with a high population density, the highest urban wage but quite high urban-rural real wage inequality. Together with Dongguan they share the highest capital/population ratio. Proportionally to the population, the capital Guangzhou, Zhuhai and especially Shenzhen have a low prevalence of single females and they also have a high late marriage rate, indicating both a better educated and slightly older population. At the other extreme, cities 14–16 have low degrees of urbanization and relatively low real urban wages and real city income although the rural wage is not very low. The highest inward net migration is into the Pearl River cities 1–3 and Huizhou and Foshan. The city with the highest outward net migration is Shaoguan.

Table 3 (Table 3 Inequality between cities over time (coefficient of variation)) shows the coefficient of variation across cities of the variable in question through time. It indicates growing inequality between the cities over time in both rural and urban wages and in capital stock per inhabitant. Interestingly variations in the late marriage rate between cities is falling but variations in the number of single females between cities and size differences between cities are roughly constant. Urbanization seems to be spreading slowly across cities so that variations between cities are gently falling. In a word overall, the rapid development since 1995 has generally been accompanied by an increase in inequality between city areas.

4.1 Applying the Theory to Guangdong

Guided by earlier studies and the descriptive statistics above, we select four factors as motives for migration:

- (a) to work as an employee in the three chief employing organizations⁷ in j , in which case the primary motivation is an expected real wage difference between cities.
- (b) to be self-employed in j , where the difference in profit opportunities between i and j accounts for the move. The size of the market P is one proxy measure, capital

Table 1 Average value of key variables

City	NM/P	Real Wu (yuan)	Real Wr (yuan)	Urban Hukou/P	Late marriage rate(%)	K/P(1000 yuan)	Single female (million)	Population (million)	City size (million m squ)
Guangzhou1	0.0056	1.49	0.84	0.61	78.51	7.11	1.90	6.4	7.26
Shenzhen2	0.0252	1.62	0.62	0.74	82.21	36.55	0.22	0.94	2.05
Zhuhai3	0.0150	1.56	0.80	0.61	81.96	20.96	0.17	0.61	1.65
Shaoguan5	-0.0010	1.20	0.90	0.34	75.20	2.03	0.93	2.9	18.38
Heyuan6	0.0025	1.28	0.83	0.18	60.72	0.61	0.97	2.9	15.48
Meizhou7	-0.0003	1.29	1.05	0.17	45.20	0.78	1.39	4.6	15.93
Huizhou8	0.0064	1.31	0.81	0.30	75.67	4.47	0.81	2.5	10.66
Shanwei9	0.0036	1.40	0.93	0.23	68.16	0.60	0.78	2.5	5.17
Dongguan10	0.0015	1.34	0.62	0.25	73.78	10.27	0.47	1.4	2.47
Zhongshan11	0.0025	1.44	0.70	0.28	76.07	6.48	0.37	1.2	1.80
Jiangmen12	0.0008	1.31	0.86	0.33	79.48	3.57	1.12	3.7	9.44
Foshan13	0.0048	1.41	0.94	0.41	72.93	8.48	0.95	3.1	3.87
Yangjiang14	-0.0008	1.36	1.05	0.24	68.51	0.68	0.71	2.4	7.81
Zhanjiang15	0.0017	1.14	0.83	0.22	76.11	1.24	1.91	6	12.49
Maoming16	0.0008	1.25	1.12	0.16	64.32	0.95	1.74	5.7	11.45
Qingyuan18	-0.0002	1.24	0.74	0.18	68.91	0.93	1.17	3.6	19.02
Mergedcity22	0.0004	1.37	0.88	0.23	72.36	1.88	3.44	11	10.28
Mergedcity23	0.0007	1.37	0.96	0.24	57.39	1.59	1.82	5.9	22.84

Table 2 Description of key variables

Description of key variables	
Wu	$WU_{it} = \frac{\text{average wage index in top three units}_{i,t}}{\text{CPI index}_{i,t}}$
	where, average wage index in top three units $_{i,t} =$
	$\frac{\text{average wage in top three units}_{i,t}}{\text{average wage in top three units}_{i,t=0}}$
	and, CPI index $_{i,t} = \frac{\text{CPI}_{i,t}}{\text{CPI}_{i,t=0}}$
	$\overline{WU}_i = \sum_{t=0}^9 WU_{i,t} \quad (*)$ $i = 1..18, t = 0..9$
wr	$WR_{i,t} = \frac{\text{average gross rural income index}_{i,t}}{\text{CPI index}_{i,t}}$
	where, average gross rural income $_{i,t} =$
	$\frac{\text{gross rural primary industry output}_{i,t}}{\text{rural primary industry labor force}}$
	and, average gross rural income index $_{i,t} =$
	$\frac{\text{average gross rural income}_{i,t}}{\text{average gross rural income}_{i,t=0}}$ and, CPI index $_{i,t} = \frac{\text{CPI}_{i,t}}{\text{CPI}_{i,t=0}}$ $\overline{WR}_i = \sum_{t=0}^9 WR_{i,t} \quad (*)$ $i = 1..18, t = 0..9$
UH rate	urban hukou rate $_{i,t} = \frac{\text{urban hukou population}_{i,t}}{\text{city population}_{i,t}}$
	$\text{urban hukou rate}_i = \sum_{t=0}^9 \text{urban hukou rate}_{i,t} \quad (*)$ $i = 1..18, t = 0..9$
Lmar-rate	Late marriage rate $_{i,t}$ is the ratio of the number of females who were at least 23 years old at marriage to the total number of first marriages.
K/P	$(K/P)_{i,t} = \frac{\text{capital stock}_{i,t}}{\text{city population}_{i,t}}$
	capital stock $_{i,t} = K_{i,t} = (1 - \delta)K_{i,t-1} + FDI_{i,t}$
	where δ is depreciation rate. $\delta = \frac{\text{original}K_{i,t=1992} - \text{net}K_{i,t=2}}{\text{original}K_{i,t=2}}$ $i = 1..18, t = 0..9$
SFP	number of single female $_{i,t} =$
	number of female $_{i,t}$ – number of married and child bearing age $_{i,t}$. where, child bearing age is between 20–49 years old $i = 1..18, t = 0..9$

intensity K/P is another. These variables should reflect the demand for services of the self-employed and their cost determining variables.

- (c) to get married/join friends etc who are in j . One measure of the relative desirability of different cities in their marriage opportunities is given by the gender structure of the population of single individuals in different cities. We measure this by the number of single females.
- (d) to leave a mainly agricultural city area to move to a more urbanized city area where infrastructure is better developed, measured by the % of the urban population holding an urban hukou.

We model the city specific migration costs by a mix of a distance measure (a city specific constant A_j) and the % of the urban population holding an urban hukou. The latter reflects

Table 3 Inequality between cities over time (coefficient of variation)

Year	NM/P	Real Wu (yuan)	Real Wr (yuan)	Urban Hukou/P	Late	Single	Population (million)	
					marriage rate(%)	K/P(1000 yuan)		female (million)
1990	1.50	0.00	0.00	0.56	.	1.29	0.68	0.69
1991	.	0.04	0.10	0.56	0.22	1.36	0.68	0.67
1992	.	0.06	0.11	0.57	0.19	1.42	0.67	0.70
1993	1.85	0.09	0.18	0.54	0.18	1.45	0.68	0.68
1994	2.01	0.07	0.19	0.53	0.19	1.51	0.69	0.69
1995	1.95	0.08	0.20	0.53	0.14	1.54	0.66	0.67
1996	1.47	0.11	0.22	0.53	0.11	1.55	0.69	0.67
1997	1.64	0.13	0.24	0.51	0.10	1.58	0.69	0.70
1998	1.72	0.16	0.26	0.52	0.11	1.65	0.69	0.69
1999	1.87	0.17	0.28	0.51	0.10	1.62	0.71	0.72

the severity with which the hukou policy is applied in different areas, a high ratio indicates a greater barrier. Note we are using the urban hukou ratio to reflect two different and opposite signed effects on net migration.

Our theory as expounded in (1) determines net migration flows from the wage, self-employment profit, marriage chance and urban infrastructure gaps between cities together with the migration costs

$$\begin{aligned}
 NM_{j^*} &= A_{j^*} + \sum_{i \neq j^*} N_i (\sum_k (a_k + b_k \mu_k^i) - c_j) - N_{j^*} (\sum_k (a_k + b_k \mu_k^{j^*}) - c_{j^*}) \quad (2) \\
 NM_{j^{**}} &= A_{j^{**}} + N_{j^*} (\sum_k (a_k + b_k \mu_k^{j^{**}}) - c_{j^{**}}) - N_{j^{**}} (\sum_k (a_k + b_k \mu_k^{j^*}) - c_{j^*}) \\
 NM_i &= A_i - N_i (\sum_k (a_k + b_k \mu_k^{j^*}) - c_{j^*}) \text{ for } i \neq j^*, j^{**}
 \end{aligned}$$

This has the interpretation of a gaps model (Zhu 2002) in a multi area and multivariate context⁸. There are 6 coefficients to estimate for each city giving a total of 108 regression coefficients.

Having defined the rankings of cities⁹ such that the top ranked is the most desirable, and individuals move when possible to the top two ranking cities, the coefficients b_k should generally be positive. There are some possible ambiguities: differences in the number of single females may proxy the availability of female worker jobs especially in the textile industries (Fan 2003; Huang 2001) or may proxy marriage chances. Similarly, depending on whether capital and labour are substitutes or complements, capital stock can have an ambiguous effect on employment prospects via the demand for labour. Finally, population can have an ambiguous effect, it could reflect disadvantages due to congestion in an area or the level of demand for the services and output of the self-employed.

We add a disturbance ε_{it} which is assumed to have a zero mean at each i, t and initially for given i , to be independent over time t with a constant covariance matrix across cities. We test the lack of autocorrelation of the residuals following estimation using Wooldridge's (2002) panel serial correlation test.

Adding the disturbance and using more succinct notation, (2) becomes

$$NM_{it} = A_i + \sum_{x|i=j^*} B_i^x Gap_{it}^x + \varepsilon_{it}, E\varepsilon_{it} = 0, E\varepsilon_{it}\varepsilon_{js} = \sigma_{ij} \text{ for all } t, s$$

Table 4 Description of explanatory variables

Description of explanatory variables	
NM	Number of net migrants
Ewu-gap	If $i = 1(x)$, expected urban wage $_{1(x)} * \sum$ urban hukou $_j$
	–expected urban wage $_{2(x)} * \text{urban hukou}_{1(x)}$
	If $i = 2(x)$, expected urban wage $_{2(x)} * \text{urban hukou}_{1(x)}$
	–expected urban wage $_{1(x)} * \text{urban hukou}_{2(x)}$
	If $i = 3(x)$, -expected urban wage $_{1(x)} * \text{urban hukou}_{j(x)}$; $j = 1 \dots 18, i \neq j, 1(x)$ stands for the highest expected urban wage city
P-gap	If $i = 1(x)$, city population $_{1(x)} * \sum$ agricultural hukou $_j$
	–city population $_{2(x)} * \text{agricultural hukou}_{1(x)}$
	If $i = 2(x)$, city population $_{2(x)} * \text{agricultural hukou}_{1(x)}$
	–city population $_{1(x)} * \text{agricultural hukou}_{2(x)}$
	If $i = 3(x)$, -city population $_{1(x)} * \text{agricultural hukou}_{j(x)}$; $j = 1 \dots 18, i \neq j, 1(x)$ stands for the biggest city population city
K-gap	If $i = 1(x)$, capital stock $_{1(x)} * \sum_j$ agricultural hukou
	–capital stock $_{2(x)} * \text{agricultural hukou}_{1(x)}$
	If $i = 2(x)$, capital stock $_{2(x)} * \text{agricultural hukou}_{1(x)}$
	–capital stock $_{1(x)} * \text{agricultural hukou}_{2(x)}$
	If $i = 3(x)$, -capital stock $_{1(x)} * \text{agricultural hukou}_{j(x)}$; $j = 1 \dots 18, i \neq j, 1(x)$ stands for the largest amount of capital stock city
SFP-gap	If $i = 1(x)$, single female $_{1(x)} * \sum$ single female $_j$
	–single female $_{2(x)} * \text{single female}_{1(x)}$
	If $i = 2(x)$, single female $_{2(x)} * \text{single female}_{1(x)}$
	–capital stock $_{1(x)} * \text{single female}_{2(x)}$
	If $i = 3(x)$, -single female $_{1(x)} * \text{single female}_{j(x)}$; $j = 1 \dots 18, i \neq j, 1(x)$ stands for the smallest number of single female city
UH-gap	If $i = 1(x)$, urban hukou $_{1(x)} * \sum$ agricultural hukou $_j$
	–urban hukou $_{2(x)} * \text{agricultural hukou}_{1(x)}$
	If $i = 2(x)$, urban hukou $_{2(x)} * \text{agricultural hukou}_{1(x)}$
	–urban hukou $_{1(x)} * \text{agricultural hukou}_{2(x)}$
	If $i = 3(x)$, -urban hukou $_{1(x)} * \text{agricultural hukou}_{j(x)}$; $j = 1 \dots 18, i \neq j, 1(x)$ stands for the biggest number of urban hukou city

where (details of the gaps are defined in Table 4 Description of exalanatory variables).

$$\begin{aligned}
 \text{Gap}_{it} &= \sum_{j \neq i^*} \mu_j * N_j - \mu_{j^*} * N_{j^*} \text{ if } i = j^* \\
 &= \mu_{j^*} * N_{j^*} - \mu_j * N_j \text{ if } i = j^{**} \\
 &= -\mu_j * N_j \text{ if } i \neq j^*, j^{**}
 \end{aligned}$$

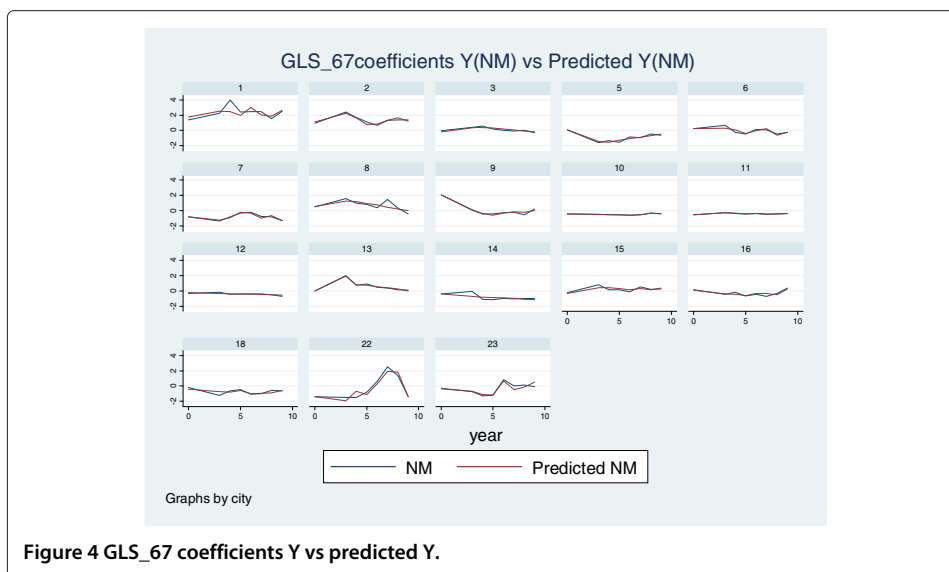
Secondly we have to specify the covariance structure between cities. For each city the variance is constant over time since it is iid. But the variances could differ between cities, effectively giving the disturbances a panel structure. Similarly the shocks of any two cities may be positively correlated (like a common global shock to all cities) or negatively correlated (eg if there is some uncertainty over the best destination for a migrant who has decided to move within Guangdong). We use the Pesaran (2006) and Hoyos and Sarafidis (2006) test for cross section dependence. The test statistic has an approximate normal distribution which should be valid even in small samples.

Table 5 GLS_67 coefficients

City	GLS_67coefficients					
	Cons	Wugap	Pgap	Kgap	SFgap	UHgap
1		0.819*			0.675*	4.109***
		2.43			2.02	4.7
2	16.98***	3.323***	(-)1.971***		(-)0.230***	19.49***
	6.52	5.66	-3.97		-4.08	6.06
3	2.317***			0.158***		
	4.24			5.39		
5	1.834***	0.313***		(-)0.421***		
	3.34	4.75		-7.39		
6	5.570***	(-)0.862*	(-)1.614**	0.160*		5.690***
	4.31	(-)2.49	(-)2.66	2.24		3.39
7		0.642***	3.293***	(-)0.379***		(-)5.461***
		3.74	8.81	(-)5.37		(-)7.07
8	2.592***			0.299**		
	4.64			2.73		
9	1.570*	1.143***	(-)1.463**	(-)0.486***	0.224*	
	2.53	7.87	(-)3.12	(-)9.19	2.26	
10	3.333***	(-)0.718***		0.212***	(-)0.130***	1.901***
	5.86	(-)8.34		7.48	(-)9.32	8.53
11	2.702***	(-)0.290**	(-)0.312**	0.109***		1.136**
	4.47	(-)3.04	(-)3.23	5.84		
12	2.115***	0.0689*				
	3.87	2.36				
13	1.190*	(-)0.743***		0.617***		(-)0.891***
	2.14	(-)12.55		19.45		(-)8.02
14	2.747***					0.790*
	4.21					2.26
15	1.707**		(-)0.656**	0.185**		
	2.96		(-)2.9	2.61		
16		0.7478***		(-)0.340***	0.352**	(-)2.792***
		4.09		(-)4.72	2.84	(-)3.77
18	2.626***		0.886*		(-)0.250*	
	4.16		2.25		(-)2.28	
22	6.549***	(-)2.465***	7.525***	(-)0.983***	0.865*	14.52***
	6.22	(-)7.76	3.4	(-)5.08	2.37	6.68
23		0.226		(-)0.296**	0.813***	(-)2.785***
		1.84		(-)2.86	4.71	(-)3.77

*p<0.05, **p<0.01, ***p<0.001, (-)negative.
t statistics are presented below coefficients.

We estimate the parameters by GLS allowing for the variances of disturbances to differ by city. In order to check the robustness of GLS, we also estimate by weighted OLS and find equivalent results. We allow the constant terms (A_i) and all the slope coefficients B_i^x to vary by city through the use of dummy variables for each city. The most general model has 108 regression parameters¹⁰, which shows no evidence of serial correlation or cross section dependence and also no evidence of panel effects or heteroscedasticity in the



disturbances. So subsequently we use a diagonal covariance matrix of the form $E\varepsilon_{it}\varepsilon_{js} = 0$ for all t, s and $i \neq j$ but $E\varepsilon_{it}^2 = \sigma_i^2$ for all t .

We impose zero and equality restrictions on coefficients to reduce the system to 67 coefficients (Table 5 GLS_67 coefficients). The restrictions are accepted on a likelihood ratio test, applied sequentially and still has no autocorrelation or cross city correlation (Table 6 Diagnostic test statistics). Estimating the same model by weighted OLS (allowing for disturbance variances to vary by city) gives very similar coefficients, an $R^2 = .959$ and easily passes a Ramsey Reset test. The weighted OLS residuals also show no sign of autocorrelation or cross section dependence. The evidence is that this base model with 67 coefficients is an adequate specification of the migration process. The accompanying plots show the relation between the actual and predicted net migration by city (Figure 4 GLS_67 coefficients Y vs predicted Y). Generally the model is replicating the data as one would expect, smoothing some of the sharper fluctuations especially in cities 1, 8, 22.

However many of the coefficients are very similar across cities, although there are some outlying gap-city combinations. Testing for equality of coefficients to group city and factor effects leads to a final model with just 27 coefficients (Table 7 GLS_27 coefficients) with a loglikelihood of -29.60 and the model also passes all the diagnostic tests. Comparing the plots (Figure 4 GLS_67 coefficients Y vs predicted Y and Figure 5 GLS_27 coefficients Y vs predicted Y) of the actual and predicted values by city for the 27 and 67 coefficient models reveals that we lose relatively little in terms of goodness of fit from imposing these restrictions on the 67 coefficient base model. We still track the data quite well and pick up most turning points in the net migration data. The 27 coefficient model represents our final model. For the sake of robustness we also estimate the final model by weighted least squares with weights being the estimated standard deviations of residuals for each city (so there is cross section heteroscedasticity but no cross section dependence). We include a constant term to allow the conventional calculation of R^2 (which is 841). The coefficient estimates and standard errors are very similar for weighted least squares and panel based GLS, and the weighted least squares results also satisfy diagnostic

Table 6 Diagnostic test statistics

	GLS	GLS	WLS	GLS	WLS
Coefficients	108	67	67	27	27
Wooldridge SR	F(1, 17) = 4.097, p = 0.59		F(1, 17) = 2.531, p = 0.13		F(1, 17) = 3.55, p = 0.08
Pesaran test	N(0, 1) = -1.12, p = 0.261		N(0, 1) = -0.70, p = 0.486		N(0, 1) = -0.59, p = 0.55
R ²			0.959	0.841	
Reset test			F(3, 73) = 0.33, p = 0.80		F(3, 113) = 0.64, p = 0.59
Sargan test					p = 0.265
Hausman-Wu test					p = 0.42

Table 7 GLS_27coefficients

	GLS_27coefficients	
NM(person)		
Real urban expected urban wage gap(yuan)	1.133***	(4.15)
Population gap(person)	-0.764***	(-10.03)
City1&2&2&23 Singel female gap(Person)	1.207***	(5.13)
Urbanization level gap(urbanhukou)	1.769***	(11.68)
City2 constant effect	4.973***	(4.49)
City2 urbanization level gap(urbanhukou)	5.516***	(4.16)
City3 constant effect	0.174***	(7.62)
City5 constant effect	-0.107**	(-2.60)
City7 population gap(person)	2.110***	(3.48)
City7 urbanization level gap(urbanhukou)	-2.122***	(-3.52)
City8 constant effect	0.562***	(4.96)
City8 Capital stock gap(yuan)	0.298**	(2.76)
City9 constant effect	0.374***	(5.67)
City9 Real urban expected urban wage gap(yuan)	0.767***	(10.17)
City9 Capital stock gap(yuan)	0.470***	(6.99)
City11 constant effect	0.0317**	(3.22)
City13 Capital stock gap(yuan)	0.161	(1.58)
City13 urbanization level gap(urbanhukou)	-0.351**	(-3.22)
City14 urbanization level gap(urbanhukou)	0.178***	(6.35)
City16 urbanization level gap(urbanhukou)	0.315***	(7.74)
City18 constant effect	-0.146***	(-5.12)
City22 constant effect	4.448***	(5.00)
City22 real urban expected urban wage gap(yuan)	-2.438***	(-7.62)
City22 population gap(person)	8.206***	(4.70)
City22 Capital stock gap(yuan)	-1.022***	(-5.97)
City22 urbanization level gap(urbanhukou)	14.94***	(7.68)
City23 Capital stock gap(yuan)	-0.440***	(-4.28)
Observations	144	

t statistics are in parentheses.

* p < 0.05.

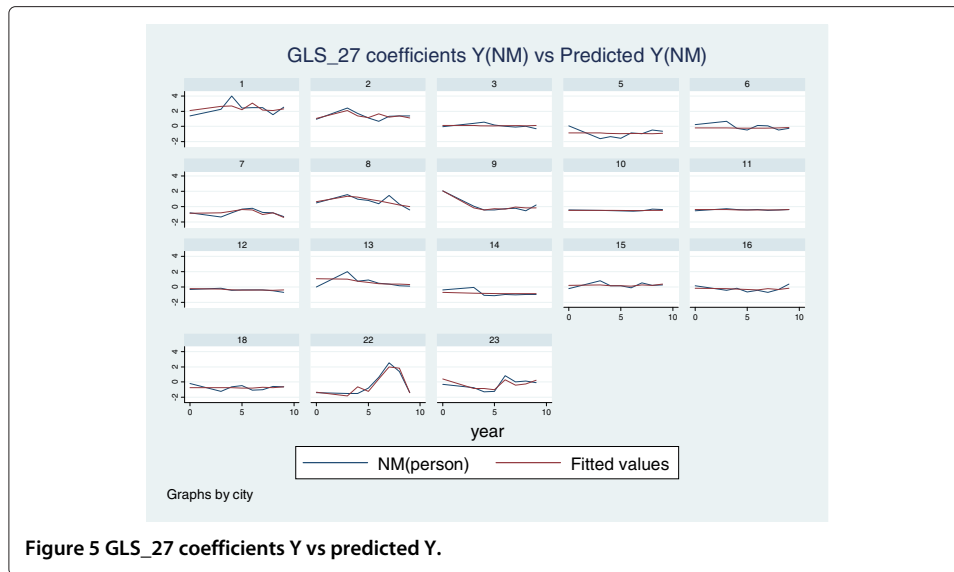
** p < 0.01.

*** p < 0.001.

tests. With Anderson's small sample log likelihood correction, the restrictions in the final model are also accepted against both the initial 108 and the base 67 coefficient model.

There is an argument that there may be some endogeneity in the regressors especially in the wage gap variable. Shocks in net migration may feed back through the city labour market into shocks in the real wage. Thus the wage gap variables may be correlated with the net migration disturbances. We instrument the three wage gap variables by FDI and employment for the common group of cities and for cities 9 and 22, giving 6 instruments in all (the Sargan test for overidentification has a p value of 0.265) and perform a Hausman-Wu test of the difference between the IV and the OLS estimates. It is not significant (the p value is 0.42), and so we conclude that there are no significant feedback effects between net city migration and the city wage gap variable.

In this reduced model all cities except for 9 and 22 have a common positive effect of expected wage differences. The expected wage gap does affect net migration into cities 9,



22 but to a smaller extent than in the other cities. Only a few cities have a responsiveness of net migration to the level of capital stock (cities 8, 9, 13, 22 and 23) and in these cities there are heterogeneous reactions to capital stock. The population gap is important in affecting net migration in all cities but only the effects in cities 7 and 22 are positive and heterogeneous whereas in the other 16 cities the response to the gap is negative although quantitatively small. All cities have net migration effects of their degree of urbanization, this is an equal effect in 12 cities but there are heterogeneous effects in 6 cities (cities 2, 7, 13, 14, 16 and 22). The effect of the gaps is thus common for most cities and most gaps.

The city specific constant terms in Table 4 reflect a relatively constant stream of net migration which is due to unobservable or non-measured city amenities (Davies et al. 2001). These factors are not determined by the operation of the gaps. These effects are important in half of the cities and in the majority of these cities there is inward migration which is not related to the gaps that we have identified.

Generally the gaps work in a way that is consistent with the theory¹¹: the population gap is a broad exception but its role generally is dominated by the merged city 22 which is very much larger than the other cities in terms of population. There are some other specific exceptions like the negative impact of the urbanization gap on net migration into city 7, indicating that, for that city, the hukou migration cost element outweighs the benefits of urbanization. Most of the heterogeneous gap effects can be explained in terms of special city characteristics. City 2 has been one of the fastest growing cities in terms of capital stock and net migration. It is a high urban wage, densely populated and highly urbanized location. City 7 is a northern mountainous city with low urban wage and urbanization, high population but low population density and low capital stock. It shows high emigration. City 8 has a relatively high capital stock and low population and its textile industry base does not yield very high expected urban income, nevertheless it attracts immigration. City 9 is a coastal city and is the main Guangdong seafood producer with other industry concentrated on shipping construction. It is a low population and low population density city, with a low expected urban income and low capital stock, but despite this it has mean positive net migration. Foshan (13) is one of the industrial tigers with high expected

urban income, capital stock and population and a relatively high degree of urbanization. It attracts positive net migration but is neither the leading nor second city in terms of the gap rankings. Cities 14 and 16 are low expected urban wage, relatively rural cities with low capital stock and average or low population density. Their mean net migration is close to zero. The merged city 22 stands out as having the greatest amount of specific heterogeneity in the migration response to gaps. As stated above it dominates the other cities in population size but is relatively non-urbanised although it has a high population density. It also has low capital stock and at best average expected urban income. It's mean net migration is close to zero. Finally city 23 is a similar administratively merged city, sharing many of the characteristics of city 22.

5 Conclusions

Individuals are in heterogeneous circumstances and any one individual is affected by many different utility relevant variables. If individuals can locate in different possible places which also have heterogeneous characteristics, then we would expect movement of individuals between locations. We develop such a multi-motive and multi-location theory to determine the aggregate net migration flows between locations. We add the assumption that all individuals have identical preferences defined over multiple location specific characteristics and, at a general level, derive the result that there will be a tendency to agglomeration. All individuals will agree on a ranking of locations from best to worst. Those individuals with bad experiences in their current location will gain the most by moving to the location which is universally judged the best. Individuals who have current utility above the average for their present location may prefer to remain in situ especially if the migration cost is substantial. So less attractive locations will have emigration especially of the lowest utility inhabitants while the best location will have inward movement. Individuals who start off in the best location, but whose individual experience in that location is much below the location average may find it advantageous to move into the second most attractive location. So on balance the second best location may have net immigration or net emigration. We would expect agglomeration into the top two cities to occur. This matches up with some of the settlement patterns predicted in economic geography type models. There are some theoretical innovations in migration modelling. Our approach allows for multivariate determinants and multi-location choices of net migration flows. People move to places where the chance of an improvement of their current circumstances in some dimension is highest. We confirm the basic Harris-Todaro insight that expected labour income differences are important but also confirm Krugman's view that each location has a variety of push and pull factors determining migration.

The framework we use is static and one period. We abstract from temporary migration, planned reverse migration and commuting/guest working. We also work with an individual as the decisionmaker which allows us to avoid specifying family decision processes and to derive aggregate net migration equations in a multi-motivation, multi-destination setting. A partial justification is that much aggregate migration data is at the household not individual level, and abstracting from intra-family decision rules yields empirically testable equations. However clearly a next step would be a multi-period and family based model.

We then use this framework to study the net migration flows between 18 different regions of the Guangdong province in China. Guangdong is particularly suitable for this

purpose since it has experienced very rapid growth and industrialization in conjunction with high levels of inward immigration from the other Chinese provinces, and even higher levels of intra-province migration. We divide Guangdong into 18 city areas which have varying degrees of urbanization and use panel data on these 18 cities for 1990–1999 to econometrically investigate net migration flows between the cities allowing for cross section heteroscedasticity. We find that net migration into the majority of cities can be well explained by a common set of parameters. There is some limited heterogeneity between cities in how net migration responds to the differentials, out of a potential total of 90 city-differential heterogeneities we find that we need just 15 specific coefficients. The remaining heterogeneities are in the impacts of capital stock and the degree of urbanization. Nearly half of the cities share a common mean amount of net migration which is unrelated to the four differentials we identified. No cross section dependence and serial correlation are detected in the final model. In terms of goodness of fit and tracking the data city by city, our model performs well and there is no evidence of model misspecification.

In a locational equilibrium, the net benefits of moving between cities should be equalized. In fact inequality between cities in some of the relevant factors has increased not fallen over our sample. The coefficient of variation of urban/rural income and capital stock per capita suggest that inequalities between cities are increasing over time. Taken together, the rising inequality in some migration inducing factors may imply that a full locational equilibrium has not yet been achieved.

It is well known that Chinese labour migration is substantial and exhibits different types of flows. It is widely argued to be a very important component in rapid Chinese growth and development, thus its policy importance is clear. Although the data sources are much more abundant than 20 years ago, there is still a paucity of degrees of freedom and coverage of some of the relevant factors. This forces some imperfection in our modelling strategy, but, given this, the results here are robust to a range of specification tests.

Endnotes

¹To our knowledge, our work is the first to aggregate migration flows from individual decisions allowing for multiple pull factors and multiple locations. Bazzi (2012) develops a microfounded model of aggregate migration flows to study to what extent financial barriers limit international migration. But he did not consider multiple pull factors and multiple location choices.

² The idea is that the factors measure variables like the wage, employment opportunities, infrastructure and local public goods, profit prospects for self-employed individuals, marriage prospects, etc.

³ In fact in this special case where the matching is one-sided, if the realised utility has the form $\varepsilon f(z) + g(z)$, where $f(\cdot) > 0$ the same argument follows.

⁴ These are an aggregation of the 21 administrative city areas to ensure unique boundaries over 1990–1999. Each city area has both an urban and rural part. The data is taken from Guangdong Statistical Yearbooks Guangdong Statistical Yearbook (1990–1999).

⁵ $K_t = (1 - \delta)K_{t-1} + FDI_t$, where δ is the depreciation rate. The base value of capital stock is given by the 1992 historic cost value of assets. The depreciation rate is computed as the % difference between the net value of fixed assets and the historic value of fixed assets in

1992, the mean of this is about 25%.

⁶ Defined as the number of females who were at least 23 years old at marriage as a proportion of the total number of first marriages.

⁷ These are either state or urban collective units, or private sector units with joint ownership, shareholding or foreign ownership (ie excluding self-employment).

⁸ Instead of thinking of the distribution of the factor within a city, this can be interpreted as saying for example that net (and gross) migration from a city ranked three or lower into the top city is the total factor gain from the lower rank city achieving the mean factor of the top city. We can also identify the constant term with a combination of a constant net migration flow unrelated to the gaps (giving a positive A_i) and a moving cost effect which deters some of the net migration driven by the gaps (giving a negative A_i).

⁹ The ranking of cities on different factors may vary with time. Broadly the Pearl delta cities (1,2,3), Foshan and city 22 are generally ranked in the top two on most factors. The implication is that we should expect to see intra-Guangdong emigration from the remaining thirteen cities which are never ranked in the top two on any criterion for migration but inward immigration into the ranked cities.

¹⁰ In fact we drop one regressor to avoid multicollinearity, leaving 107 coefficients.

¹¹ Since we have scaled the regressors to have zero mean and unit variance across the whole sample, the estimated coefficients are largely independent of the units in which we measure variables. So we do not compute distinct elasticities.

Competing interests

The IZA Journal of Migration is committed to the IZA Guiding Principles of Research Integrity. The authors declare that they have observed these principles.

Acknowledgements

Helpful suggestions from a referee are gratefully acknowledged. Errors and omissions are our responsibility. Responsible editor: Denis Fougère

Received: 26 October 2012 Accepted: 17 March 2013

Published: 29 April 2013

References

- Bazzi S (2012) Wealth heterogeneity, income shocks, and international migration: theory and evidence from Indonesia. [online] Available at: <http://ipl.econ.duke.edu/bread/papers/1012conf/Bazzi.pdf>
- Borjas GJ (1999) The economic analysis of immigration, Handbook of labor economics. Harvard University
- Brucker H, Jahn E (2008) Migration and the wage curve: a structural approach to measure the wage and employment, IZA working paper series. [online] Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1294546
- Chan KW, Zhang L (1999) The hukou system and rural-urban migration in China: processes and changes. *China Q* 160: 818–855
- Charsley K, Liversage A (2011) Transforming polygamy: migration, transnationalism and multiple marriages among Muslim minorities. In: Rogers A, Vertovec S, Cohen R (eds). *Global Networks*. Blackwell Publishing Ltd & Global networks partnership, UK
- Chen Y, Rosenthal Stuart S (2008) Local amenities and life-cycle migration: Do people move for jobs or fun? *J Urban Econ* 64: 519–537
- Davies PS, Greenwood MJ, Li H (2001) A conditional logit approach to U.S. state-to-state migration. *J Reg Sci* 2: 337–360
- Docquier F, Rapoport H (2005) The economic of migrants' remittances. In: Kolm SC, Ythier JM (eds). *Handbook on the economic giving, altruism and reciprocity*, 1135–1198. Elsevier, North-Holland. Ch. 17., [online] Available at: <http://www.sciencedirect.com/science/article/pii/S1574071406020173>
- Dustmann C, Frattini T, Glitz A (2007) The impact of migration: a review of the economic evidence, final report No.102. UCL: Centre for Research and Analysis of Migration (CReAM). [online] Available at: http://www.ucl.ac.uk/~uctpb21/reports/WA_Final_Final.pdf
- Ekeland I, Heckman JJ, Nesheim L (2004) Identification and estimation of hedonic models. *J Pol Econ* 112(1): s60—s109
- Fan C (1996) Economic opportunities & internal migration: a case study of Guangdong. *Prof Geographer* 48(1): 28–45
- Fan, C (1999) Migration in a socialist transitional economy: heterogeneity, socioeconomic and spatial characteristics of migrants in China and Guangdong Province. *Int Migr Rev* 33(4): 954–987
- Fan C (2002) Marriage & migration in transitional China: a case study of Gaozhou, Western Guangdong. *Environ Plann A* 34: 619–638
- Fan, C (2003) Rural-urban migration and gender division of labour in transitional China. *Int J Urban Reg Res* 27: 24–27

- Fan CC, Huang YQ (1998) Waves of rural brides: female marriage migration in China. *Annu Assoc Am Geographers* 88(2): 227–251
- Federici D, Giannetti M (2010) Temporary migration and foreign direct investment. *Open Econ Rev* 21: 293–308
- Furtado D, Theodoropoulos N (2008) I'll marry you if you get me a job: cross-nativity marriage and immigrant employment rates, CReAM discussion Paper. CDP No 01/08. [online] Available at: http://www.cream-migration.org/publ_uploads/CDP_01_08.pdf
- Greenwood MJ (1997) Internal migration in developed countries. *Handbook of Population and Family Economics*, v.1, 47–720. Elsevier, North-Holland. Ch. 12. [online] Available at: <http://www.sciencedirect.com/science/article/pii/S1574003X97800049>
- Greenwood MJ, Hunt GL, Mcdowell JM (1986) Migration and employment change: empirical evidence on the spatial and temporary dimensions of the linkage. *J Reg Sci* V(26): 223–234
- Harris JR, Todaro MP (1970) Migration, unemployment and development : a two-sector analysis. *Am Econ Rev* 60: 126–124
- Heckman JJ, Matzkin R, Neshheim L (2002) Identification and estimation of hedonic models: the vector nonseparable case with missing attributes. Univ. Chicago, Dept, Econ, Chicago
- Hoyos DE, Sarafidis V (2006) Testing for cross-sectional dependence in panel-data models. *Stata J* 6: 482–496
- Huang Y (2001) Gender, hukou and occupational attainment of female migrants in China. *Environ Plan* 33: 257–279
- Johnson DG (2003) Provincial migration in China in the 1990s. *China Econ Rev* 14: 22–31
- Krugman P (1992) A dynamic spatial model. Working Paper No 4219. National Bureau of Economic Research. [online] Available at: <http://time.dufe.edu.cn/spti/article/krugman/krugman012.pdf>
- Kugler M, Rapoport H (2005) Skilled emigration, business networks, and foreign direct investment. CESIFO W.P.n: 1455. http://www.econstor.eu/dspace/bitstream/10419/18819/1/cesifo1_wp1455.pdf
- Lilleor HB, Broeck KVD (2011) Economic drivers of migration and climate change in LDCs. *Migr Glob Environ Change* 21: s70—s81
- Mesnard A (2000) Temporary migration and capital market imperfections. *Oxford Econ Papers* 56: 242–262
- Meyerhoefer CD, Chen CJ (2011) The effect of parental labor migration on children's educational progress in rural China. *Rev Econ Household* 9: 379–396
- Nishitateno S (1983) China's special economic zones: experimental units for economic reform. *Int Comp Law Q* 32: 175–185
- OECD (2010) Entrepreneurship and migration. [online] Available at: <http://www.oecd.org/industry/smes/45068866.pdf>
- Pesaran H (2006) Estimation & Inference in large heterogeneous panels with multifactor error structure. *Econometrica* 74(4): 967–1012
- Piper N (2005) A paper prepared for the policy analysis and research programme of the global commission on international migration. GCIM. http://www.iom.int/jahia/webdav/site/myjahiasite/shared/shared/mainsite/policy_and_research/gcim/tp/TP7.pdf
- Plantinga A, Dessendre CD, Hunt GL, Piguat V (2012) Housing prices and inter-urban migration. *Reg Sci Urban Econ* 43: 296–306. <http://www.sciencedirect.com/science/article/pii/S0166046212000683>
- Poncet S (2006) Provincial migration dynamics in China: Borders, costs and econometric motivations. *Reg Sci Urban Econ* 36: 385–398
- Renard M, Xu Z, Zhu N (2011) Migration, urban population growth & regional disparity in China, CERDI , Etudes et Documents, E 2007.30. [online] Available at, <http://hal.archives-ouvertes.fr/docs/00/55/69/81/PDF/2007.30.pdf>
- RUMICI (2007) Rural-urban migrant project census manual. [online] Available at: http://rse.anu.edu.au/rumici/pdf/Census%20manual_China_English08.pdf
- Seeborg MC, Jin Z, Zhu Y (2000) The new rural-urban labor mobility in China: causes and implications. *J Socio-Econ* 29: 39–56
- Statistical Bureau of Guangdong Province (1990–1999) *Guangdong Statistical Yearbook*. 1st ed. China Statistics Press, China
- Taylor JE, Martin PL (2001) Human capital: migration and rural population change. *Handb Agric Econ* 1: 457–511
- The Migration Observatory (2012) Migrants in the UK labour markets: an overview. [online] Available at <http://www.migrationobservatory.ox.ac.uk/briefings/migrants-uk-labour-market-overview>
- Todaro M (1969) A model of labour migration and urban unemployment in less developed countries. *Am Econ Rev* 59(1): 138–148
- Vermeulen W, Ommeren JV (2009) Does land use planning shape regional economics? A simultaneous analysis of housing supply, internal migration and local employment growth in the Netherlands. *J Housing Econ* 18: 294–310
- Whalley J, Zhang S (2007) A numerical simulation analysis of (Hukou) labor mobility restrictions in China. *J Dev Econ* 83: 392–410
- Wooldridge JM (2002) *Econometric analysis of cross section and panel data*. MIT Press, Cambridge
- Zhao Y (1997) Labor migration and returns to rural education in China. *AM J Agr Econ* 79: 1278–1287. [online] Available at: <http://ajae.oxfordjournals.org/content/79/4/1278.short>
- Zhu N (2002) The impacts of income gaps on migration decisions in China. *China Econ Rev* 13: 213–230
- Zhu N, Batisse C, Li Q (2009) The flow of “peasant-workers” in China since the economic reform :a longitudinal and spatial analysis. [online] Available at: http://www.cerdi.org/uploads/sfCmsContent/html/317/Zhu_Batisse.pdf

doi:10.1186/2193-9039-2-7

Cite this article as: Simmons and Xie: **Where is the grass greener? A micro-founded model of migration with application to Guangdong.** *IZA Journal of Migration* 2013 **2**:7.