TAMPERE ECONOMIC WORKING PAPERS
NET SERIES

MIGRATION, POLICY, AND FORMATION OF CITIES

Hannu Laurila

Working Paper 28
December 2003
http://tampub.uta.fi/econet/wp28-2003.pdf

DEPARTMENT OF ECONOMICS

FIN-33014 UNIVERSITY OF TAMPERE, FINLAND

ISSN 1458-1191
ISBN 951-44-5874-5

MIGRATION, POLICY, AND FORMATION OF CITIES

5.5.2004

Hannu Laurila
Department of Economics
FI-33014 University of Tampere

Abstract:

Efficient allocation of resources is the cornerstone of market economy, and migration of people is an elementary part of the allocation mechanism. This paper presents a simple club theoretic model of migration between cities. In the model, welfare in a city depends on its size, and people migrate between cities in seek for welfare gains. Migration alone can produce stable and efficient market solutions only under very special circumstances. In general, the mechanism must be supplemented by collective optimization of population by local policy. In the most realistic case of heterogeneous cities centralized policy interventions are necessary to secure efficiency. The results about the effects of centralized policy somewhat contradict the conventional wisdom. First, administrative and economic policy measures (quantity restrictions and lump sum transfers, respectively) differ in their effects, and second, lump sum transfers actually motor up rather than stabilize spatial evolution.

Key words: agglomeration economies, city size, migration
JEL Classification: 931

1 Introduction

Efficient allocation of resources is the cornerstone of market economy. The abstract concept of resource allocation becomes more concrete once one accepts the fact that economic resources are always spatially attached to people, firms and all kinds of physical and social structures. Therefore, any reallocation of resources inevitably implies physical relocation in geographical space. The national geography in turn is composed of a set of different kinds of agglomerations of people and economic functions.

Location decisions of firms and people constitute the basic mechanism that forms the spatial economy and the cities. Location in general and urban location in particular is a central determinant of firms' profits and people's welfare. On one hand, this is because of the simple fact that the everyday transactions in the factor and goods markets take space and necessitate transit. Therefore, proximity of economic and other functions affects the transaction costs. On the other hand, there are economics of scale and scope, economies of localization and urbanization, and other externalities that have direct and indirect effects on profits and welfare. The economies of agglomeration have grown more and more important, and urban areas and cities indisputably offer the most fertile soil for the flourishing of the modern market economy.

This paper examines the working of the market mechanism in producing an efficient system of cities. Being aware of the well-known weaknesses of standard-type competitive models in the spatial context, a simple club theoretic model is constructed to provide a general equilibrium treatment of spatial evolution of the economy. To grasp the essentials about the simultaneous nature of the development, the location decisions of firms are taken to be subordinate to people's choices in that the firms are assumed to relocate instantly according to any changes in commodity and factor markets so that they are always in their profit maximizing location. Focusing on migration of people enables a simple translation of efficiency into terms of social welfare. This translation reflects the main feature of the market economy that the production factors, solely owned by the people, must be made efficiently use of so as to maximise people's utility from consumption of the final goods.

The present model is based on two main components. First, the size of a city - or proximity of people and various economic and non-economic functions - matters to the welfare it can create to its residents. Welfare in a city is a product of private and collective consumption and production activities. Most of these activities are local in nature and take place in local goods and factor markets. Due to externalities and economies of scale in production and consumption, the size of the local marketplace is relevant as to the attainable local welfare. As welfare depends on economic and non-economic locality-dependent characters, an increase of the size of a city yields benefits and cost savings, but eventually it may bring up also falling benefits and increasing costs. The collective nature of cities arises from the agglomeration economies attached with production and consumption of private goods, and from indivisibility of local public goods and services. Without these characteristics, a totally dispersed spatial pattern would sustain.

Second, in a decentralized market economy with autonomous local governments, people exert their utility maximising choices in two ways, in a purely private fashion through migration, and in a collective fashion through local democracy. By migration people seek for optimal location in order to maximise welfare. The choices are made against private calculations of welfare in different locations. Individual agents take the local circumstances as parametric, but do not take into account the external effects they cause to other individual agents. Free mobility of people based on perfect information about the benefits and cost attached to different cities creates a basic market mechanism in the spatial system.

The major flaw of free mobility as a market mechanism is that peoples' location decisions are not continuous in nature. This non-convexity problem is to say that people cannot choose optimal amounts, that is city size, like they do in conventional marketplaces. But, since the pooling up of the positive and negative effects of agglomeration depends on city size, size must be an endogenous choice variable in the model. The collective type of decision-making solves the problem: people exert their choice on city size as a collective, through local democracy. Autonomous local governments must thus as active market agents optimise on the size of the city. Free mobility and collective optimization of city size can together provide an efficient market mechanism, at least if there are no externalities or spillovers that work between the cities.

The paper proceeds as follows. Chapter 2 starts with a textbook case of migration between big and homogenous cities and illustrates the welfare effects of collective optimization on population. Chapter 3 examines the reverse case of small cities. Chapter 4 introduces the complications caused by various kinds of asymmetries in the set-up, and investigates the need, instruments and effects of centralized policy. Chapter 5 concludes the findings.

2 Big homogenous cities

2.1 Migration

The standard textbook case concerns an economy that consists of a multiple of big homogenous cities. Assume that there is no rural sector in the economy, and that there is no geographic, climatic or other variation in the national territory such that would affect the welfare creating potentials of the cities. Assume also identical technology and technical efficiency in the local private and public sectors in all cities, and that the geographical areas of the cities correspond to their long-term optima.

The first step is to examine the pure market solution of spatial allocation of population in the economy. The market mechanism is based on perfect mobility of utility maximizing people. People make the most preferred choices among the alternatives. Call this kind of a choice the exit decision (Bailey, 1999). For simplicity, fix the total population in the economy to $\underline{n}$ and the number of cities to $\underline{m}$. This is to say that the total population must always be allocated into the existing cities.

Welfare in a city is taken to depend on its size because of agglomeration economies (Richardson, 1973). The relevant measure for welfare is that experienced by individual migrants. In the presence of agglomeration economies and eventual agglomeration diseconomies, the welfare that a city can offer to its residents can be presented by an inverted U-shaped average welfare curve. It must be noted, though, that the full shape of the curve is unobservable to the people – the only operational observation is the value of average welfare that occurs at a certain point on the curve. Moreover, the average welfare curve is accompanied by a marginal welfare curve, which remains totally abstract to the people. Figure 1 presents these kinds of schedules in a pair-wise setting for two optional cities superimposed on each other.

(Figure 1 here)

In Figure 1, there are two optional sites of residence, one particular city A and city B, which represents the rest of the economy. The average welfare curve for city A, denoted by W_A, is drawn from left to right. The marginal welfare curve MW_A strikes through W_A from above at its culmination point. The respective curves for city B, denoted by W_B and MW_B, are drawn from right to left. The latter schedules represent all the remaining $\underline{m}$-1 identical cities in terms of one such city. Denote the length of the horizontal axis by N. A simpler way to deal with the setting would be to assume the economy to consist of only two cities and let $N = \underline{n}$. This interpretation would suffice in the present simple case. In any case, since all cities are alike, the two sets of curves are identical mirror images.

By strict interpretation, the use of average welfare schedules as the relevant information for individual people suggests that people should (ex ante) be assumed homogeneous. A less strict interpretation is that the schedules reflect (ex post) average or systematic responses of people. Some people may migrate against the main stream, but that kind of behaviour can be (ex post) regarded as purely stochastic. The latter interpretation is particularly tempting in the present setting, where migrants choose between homogenous alternatives, and where the systematic direction of migration is what matters.

To investigate the mechanism of free migration driven by individual welfare maximisation, suppose that there initially exists an observed welfare differential between the cities. If the initial allocation of people is described, say, by n^1, there is a welfare differential that amounts to $W_A^c - W_B^d > 0$. The attainable welfare gain attracts individual people to city A, which makes n shift rightwards in Figure 1. As a consequence, average welfare in city A falls because of increased agglomeration diseconomies, and average welfare in city B rises because of decreased agglomeration diseconomies. Migration towards city A will continue as long as it is regarded beneficial, that is until the average welfare in both cities becomes equalized. The market equilibrium occurs at the intersection point e, where

$$W_A^e = W_B^e. \tag{1}$$

The solution is stable since nobody can gain by moving into another city. As a result, n^e people will reside in city A and $(N - n^e)$ people will reside in city B. In the present case of homogenous cities the optimal population equals in every city, that is

$$n^e = (N - n^e) = \underline{n}/\underline{m}. \tag{2}$$

Efficiency of the market solution can be evaluated in terms of social welfare. The basic measure of social welfare is given by the sum of the areas below the marginal welfare curves. Because of perfect homogeneity of the cities, the marginal welfare curves MW_A and MW_B intersect in point E, which lays vertically below the equilibrium point e at n^e. Social welfare in this market solution is given by the sum of the areas O_AaEn^e and n^eEbO_B in Figure 1. Compare the sum to that in the initial allocation of population n^1 (or any allocation other than n^e) and find that the market solution is superior. Alternatively, social welfare can be measured by the sum of the areas n^e times W_A^e, and $(N - \underline{n}^e)$ times W_B^e. To include also the welfares of those people in the remaining $\underline{m}$-2 cities that are not present in the figure, the respective welfare areas have to be accounted for. In this simple case social welfare, denoted by SW, thus reads $SW^e = n^eW_A^e+(\underline{n} - \underline{n}^e)W_B^e$, which, using (1) and (2), yields

$$SW^e = \underline{n}W_i^e = \underline{m}n^eW_i^e,\ i=A,B. \tag{3}$$

The conclusion is that the market solution is efficient because social welfare is maximised. Therefore, in this special case, individual maximisation of personal welfare ends up to a social optimum that gives the maximal social welfare attainable for all the $\underline{n}$ people in the $\underline{m}$ localities.

2.2 Policy considerations

The second step is to take into account that people exert their choices not only through migration (exit) between the cities but also collectively within the cities. The concept voice refers to the collective mode of people's decision-making (Bailey, 1999). This is to say that the collective power of the cities joins the purely individualistic power of migration in the mechanism that determines the formation of cities in the economy. The non-convexity problem of exit (Stiglitz, 1977, p. 275) is overcome. Assume that the collective preference revealing mechanism within the cities is efficient.

In optimization of their population, the cities can take a total economy approach and aim to maximisation of the total welfare in the city, or they can take the within-club approach and aim to maximisation of the average welfare in the city (Cornes & Sandler, 1986, p. 175-176). Note that the term ‘total economy’ is somewhat confusing here because it refers to a single city, not to the whole economy consisting of multiple cities. By Figure 1, however, it is evident that the city-wise total economy approach is useless. The population distribution given by n^1 actually is such that maximizes total welfare in city B. This is because the MW_B curve strikes through the horizontal axis at n^1. But, as it was seen above, the situation is not sustainable, because people are motivated to emigrate from city B to city A. Local policy based on the total economy type policy rule is thus nullified by free migration.

The policy rule of the within-club approach, on the other hand, is sustainable. By setting its population so as to maximize its average welfare and by closing doors before additional immigrants, a city can shelter itself from the negative effects of migration. Implementation of this kind of a policy may, however, be difficult unless the initial starting point of migration is on the rising part of the W curve in the considered city. If so, the policy is operational. A situation depicted in Figure 1 is more academic in nature, because implementation of the policy in that case would necessitate eviction of excess people, which sounds incompatible with the concept of free market economy. Anyway, provided that the policy is fully operational, the outcome is stable since nobody can migrate after the implementation of the policy.

The outcome is not efficient in this partial equilibrium setting, however, because social welfare inevitably remains lower than that in the market equilibrium.

Examination of efficiency of collective action from the long-term general equilibrium point of view necessitates some amendments to the assumptions. If all cities apply the within-club type policy rule and constrain their size according to the culmination point of their average welfare curve, the inevitable result is that the sizes of the existing cities fall and, consequently, all the $\underline{n}$ people cannot fit into the $\underline{m}$ cities. Therefore, social welfare would fall. This problem can be overcome by letting the number of localities m to be variable. In the longer term, new cities may enter into the city system, which is to say that the number of cities becomes an endogenous variable in the model.

Assume that m is variable and let the cities optimize on their population and limit their size to the average welfare maximising level. Assume also that the general conditions concerning technology, tastes etc. remain unchanged during the adjustment period. Those people who are excluded from the set of optimal cities must found new growing agglomerations, which eventually reach their optimal size, and so on. The number of cities increases until a general equilibrium is reached. Figure 2 illustrates the outcome.

(Figure 2 here)

Figure 2 shows that, since the length of the horizontal axis in the two-city presentation is, by expression (2), $N = \underline{n}2/\underline{m}$, an increase in m will inevitably make the vertical axis shorter for the two particular cities. This in turn brings the two sets of curves closer to each other. Exclusion of people from the initial cities and founding of new homogenous cities will continue until the maximum points of the average welfare curves coincide in the pair-wise setting of Figure 2. At the equilibrium point e^* also the marginal welfare schedules intersect. Ignore the integer problem and assume that all people are included in the resulting m^* optimal cities so that

$$m^* n^{e^*} = \underline{n}. \qquad (2')$$

The result is that the welfare of all residents in all cities is maximized and equal. Moreover, in terms of the long-term equilibrium of Figure 2, $W_A = MW_A = W_B = MW_B$.

In the general equilibrium of Figure 2, total population is allocated into a set of cities that all maximise the welfare of their residents. Then also the social welfare in the economy is maximised. Social welfare can be measured by the sum of the areas below the intersecting MW curves, or by the sum of the products of population and average welfare at n^{e^*} (including the remaining $m^*\text{-}2$ cities in both versions). More formally, recalling expression (2'),

$$SW^* = m^* n^{e^*} W_i^* = \underline{n} W_i^*, \; i=A,B. \qquad (3')$$

Any other population allocation than n^{e^*} would yield lower social welfare. Furthermore, comparison of expressions (3) and (3') shows that the long-term solution is Pareto superior to the previous case of migration between a fixed number of localities, that is $SW^* > SW^e$. This is because $W_i^* > W_i^e$, that is all people are now better off since they are getting the maximal attainable welfare in the more numerous cities in which the higher welfare is experienced. Being at the top of the average welfare curves in each city is to say that agglomeration economies are optimally utilised with reference to agglomeration diseconomies. The solution corresponds to the long-term equilibrium of the competitive commodity markets, where goods are produced and consumed at minimum average cost.

The last modification to the analysis is to consider the total population n also as a variable. This is to say that people are free to migrate not only inside the economy but also between economies. Assumption of perfect inter-economy mobility of people, however, does not change the above result. As a matter of fact, it merely makes the evolution of the system of cities towards the equilibrium more credible and gives more reason to ignore the integer problem. This is because $N = 2n/m$ says that changes in m and n affect the length of the horizontal axis of Figure 2 in opposite directions.

If both m and n are variable, it is obvious that n can either increase or decrease. On one hand, exclusion of people outside the optimized cities can result in an outflow from the system, and make n decrease while m increases. The path to the equilibrium e^* becomes more straightforward. The long-term equilibrium is in principle reachable even among the initial set of $\underline{m}$ cities with a smaller number of total population. The dismal character of this result is that social welfare falls as compared to (3') and maybe also to (3) because there are less people to experience W_i^* in fewer cities of optimal size. On the other hand, the growth of new cities may attract people from outside the system, which makes n to increase along with m. The equilibrium path becomes more winding, and the natural limits of the national territory will be the final constraint to the development. The increase in n will also increase social welfare as compared to (3), at least if the newcomers' welfare is treated equally to the original people's welfare.

To sum up, the standard case of big homogenous cities yields a comforting result: efficiency is guaranteed with or without collective decision-making in the cities. Free migration alone (exit) ends up to a stable and efficient outcome, but in the longer run free migration and collective decision-making (exit and voice) together can produce a Pareto superior outcome. This precludes endogenous determination of the number of cities m in the system. In either case, there is no need for a policy intervention from the viewpoint of the social welfare of the whole economy. In general, this modification complicates the evaluation of social welfare somewhat because the sign of the change in total population n remains ambiguous. Taking total population as endogenous may give reason to centralized policy intervention to stop emigration from the economy.

3 Small cities

3.1 Migration

The setting in the above chapter was constructed under the presumption that the cities are initially big so that the market solutions appear on the falling regimes of the average and marginal welfare schedules. Migration flows that direct from bigger towards less big cities make agglomeration diseconomies in the whole system diminish. However, this kind of a situation can be regarded only as a special case, and an obvious extension would be to analyse the opposite case, where the cities are small so that agglomeration economies dominate everywhere. Therefore, turn now to the case of initially small homogenous cities, illustrated by Figure 3.

(Figure 3 here)

Figure 3 again presents two cities, one particular city A and city B as a representative of the rest of the economy. To examine the existence of a market solution of free migration, consider point *e*, where the average welfare in both cities is equal, $W_A^e = W_B^e$. In this equi-welfare position, there are no systematic gains to be achieved by moving. Nevertheless, somebody is always on the move for purely stochastic reasons. Suppose that there occurs a stochastic migration shock from city B to city A so that the population of city A is drawn to n'. As a result, a welfare gap $W_A' - W_B' > 0$ is opened up. The welfare gap now starts to attract systematic migration to city A. This, in turn, makes the welfare gap even wider, and accelerates migration further. Point *e* is clearly not a stable solution. As a matter of fact, free migration would continue until city B becomes totally deserted. If the stochastic element of migration should have emerged towards B, then A would have been deserted. The two stable market equilibriums on the left or right vertical axis are corner solutions in nature.

The evaluation of efficiency is a bit more complicated in the case of small cities than in the case of big ones. As the setting is drawn in Figure 1, the non-stable solution *e* seems to be efficient at the marginal. Because of the assumption of homogenous cities, the *MW* curves intersect at *E* vertically above *e*, irrespective to whether the two

MW curves rise or fall around point *E*. Comparing the social welfare at n^e to that at n', the conclusion is that the stochastic deviation from n^e causes a welfare loss depicted by the area *Eab*. If the corner solutions are efficient or not as compared to point *e*, however, depends on the curvatures of the *W* and *MW* schedules nearby the vertical axis. In particular, if the *W* curves hit the vertical axis above/below $W_A{}^e = W_B{}^e$, then the corner solution is Pareto superior/inferior to that at point *e*. If the latter should be the case, then there clearly is need for some kind of a policy intervention to prevent deviations from point *e* type non-stable but at the marginal efficient solutions. (Atkinson & Stiglitz, 1982, p. 533-535.)

3.2 Policy considerations

Figure 3 reveals the possible need for policy intervention. The next step of the analysis then is to ask if the intervention should be local or central in nature. Start by investigating the effects of local policy. Assume again that the cities apply the within-club type optimization on their population, and let *m* and *n* be variable. For the fortunate city A in Figure 3 local policy-making implies that the welfare increasing immigration is stopped in the culmination point of W_A, which is to say that n^A is chosen. In this case the policy is easy to implement – it suffices to close doors at the optimum. The policy instruments include city planning, dimensioning of local public services etc. While city A is stabilized at its optimal position, migration continues between the remaining cities. Some of the fortunate cities are eventually able to reach their optimum and close doors, but some less fortunate cities continue to loose population. The result is a set of optimal cities and a set of deserted ones. In the path towards a general equilibrium solution, the number of cities *m* must fall. On the other hand, the number of total population n may change in any direction. Figure 4 presents the longer-term evolution.

(Figure 4 here)

Recalling again that the length of the horizontal axis in Figure 4 is given by $N = n2/m$, a fall in m and/or an increase in n is to say that the axis becomes longer. Since the W and MW schedules for the two cities A and B stick to the respective vertical axis, the schedule sets are drawn closer to each other in the figure. In the long-term optimum the culmination points of the W schedules coincide in e^*, where the MW schedules also intersect. The result is a stable and Pareto efficient equilibrium, where nobody can gain by moving to another city. Average welfare W^* is equalised in all existing m^* cities of n^* population. Social welfare is measured by

$$SW^* = n^*m^*\ W_i^*,\ i=A=B. \qquad (3")$$

The solution is Pareto superior to the initial e type solution and to the two possible corner solutions provided that total population does not decrease. Optimality of course necessitates that the integer problem is avoided. The possible increase in n again helps in this respect and facilitates a solution that is superior to (3") also for the whole set of initial cities $\underline{m}$. In general it cannot be ruled out that the total population might also decrease. If people move outside the system rather than between the domestic cities, the evolution towards the equilibrium will take time and social welfare will erode as compared to (3").

To sum up, the market mechanism based on exit and voice type decision-making is in principle able to provide an efficient solution in the case of small cities. Centralized policy intervention is not needed unless people choose to emigrate out of the economy rather than to move inside the economy.

4 Asymmetric settings

4.1 Migration

An elementary assumption in the above two applications of the model is that the cities are not only big or small at the same time but also homogenous in their capability to create welfare for their residents. The assumption simplifies the set-up and is general enough to grasp the essential elements behind the evolution of the city system. In practice, however, this kind of an assumption is seldom satisfied. Both components of the assumption need to be critically assessed.

In practice, any system of cities usually consists of a few big cities and a lot of smaller ones. The famous Zipf's law is astonishingly valid throughout the world, at least in the long run. The simple rank size rule version of the law says that the size of the second biggest city is half that of the biggest one, the size of the third biggest city is one-third that of the biggest and so on (McCann, 2001, p. 79-80). Quite evidently, notable migration flows occur between cities of different size. Migration flows can also occur between cities in different position as to utilization of agglomeration economies. In the rank order of the cities, the largest cities (and often the most attracting ones) may encounter dominance of agglomeration diseconomies, while the smaller ones are at the same time in the regime of dominating agglomeration economies.

Second, there are good reasons to believe that the cities are heterogeneous in their capability to create welfare. This is due to variation in geography and climate, immobility of natural resources, fixed national infrastructure and networks, transport and trade connections, industrial structure, administrative status etc. Heterogeneity of the cities results in differences in the respective sets of the W and MW schedules. One gets an immediate insight into the role of the assumption by closer inspection of Figure 1: if the sets of the W and MW schedules are not perfect mirror images, the intersection points e and E do not necessarily lay vertically one upon another. This is to say that the stable solution of free migration is not necessarily efficient even in the short-term case of big cities.

It seems more than appropriate to consider the complications caused by various kinds of asymmetries in the above model of inter-city migration. For simplicity, consider two types of cities, type A and type B, of which the A type cities have an absolute advantage in generating welfare. This is to say that the average welfare curve reaches a higher peak value in city A than in city B. Figure 5 below is drawn to illustrate not only all kinds of possible settings between different types of cities but also the possible long-term equilibrium given that m and n may be variable.

(Figure 5 here)

In figure 5, the *W* and *MW* curves are again presented from left to right for a city of type A and from right to left for a city of type B. The *W* curves are drawn to intersect in two points, *a* and *b*, which suffices to illustrate all kinds of pair-wise situations between cities A and B.

Start by analyzing the situation around point *a*, where the small but prosperous city A benefits from agglomeration economies, while the much bigger but less prosperous city B suffers from excess agglomeration diseconomies. In the intersection point *a* welfare in both cities would be equal. The point *a* is not, however, a stable market solution of free migration. To the left from point *a*, a welfare gap is opened in favour of city B, and migration from A to B draws the solution further to the left. Rightwards from point *a*, a welfare gap is opened in favour of city A, and migration draws further to the right.

Around the intersection point *b*, on the other end, the more modern but overly congested city A meets competition of migrants with a traditionally equipped small city B. Now the welfare equalizing solution at point *b* is stable. On the left side of point *b* city A attracts immigrants from city B drawing the solution rightwards. The opposite is true on the right side of point *b*. Thus, point *b* is a stable market equilibrium brought up by migration

Neither of the welfare equalizing points *a* and *b* in Figure 5 is socially optimal. It is easy to see that even the stable solution at point *b* is not efficient – any other allocation of population leftwards from n^b would clearly imply higher social welfare. The higher welfare potentials of the cities remain unutilized because people are only capable to see the existing welfare differentials. The socially optimal solution occurs in point E at n^E, where the corresponding MW curves intersect reflecting optimal utilisation of agglomeration economies in the city system. Therefore, for example in the stable equilibrium point b city A is overly crowded and city B is under-populated as compared to optimal utilization of agglomeration economies in the economy.

4.2 Policy considerations

At the socially efficient allocation of population n^E, however, there is a welfare differential $W_A^E - W_B^E > 0$ in favour of city A. Thus it seems that a socially optimal allocation of people between city A and city B can be achieved only by a policy intervention that sustains the welfare gap. This seems to be necessary even if local within-club optimization of population is assumed. If both cities A and B optimize their population according to the within-club rule and m and n settle to their optimal values and no integer problem exists, the outcome is that depicted by n^* in Figure 5. At this long-term optimum allocation n^*, there is a welfare gap $W_A^* - W_B^* > 0$ in favour of city A, but the solution still deviates from the socially optimal n^E, where the gap is $W_A^E - W_B^E > 0$. More precisely, the result of local level policy-making reads for the representative pair of cities

$$SW^* = n^* W_A^* + (N-n^*)W_B^* < n^E W_A^E + (N-n^E)W_B^E = SW^E. \qquad (4)$$

By expression (4) it is obvious that reaching to the social optimum necessitates a centralized total economy approach that includes both cities A and B. The central government has the two standard options for policy instruments, the administrative instruments and the economic instruments. The administrative instruments input the policy impulse along the horizontal axis in Figure 5, and the economic instruments input the impulse along the vertical axis in Figure 5. The conventional wisdom concerning the centralised policy intervention is that the measures should be horizontally and vertically equal in their effect, or to put it more precisely, both instruments should stabilize the status quo of the spatial structure.

The administrative policy instruments include quantity rationing that is legislative and other such policy measures that produce the optimal allocation n^E and make it sustainable. In this kind of a planning solution there exist welfare differences between the two types of localities. The economy is divided into better-off and worse-off cities, which is to say that the principle of regional equity must be ignored. Implementation of the policy means that people are allowed to move from city B to city A until the allocation n^E is reached. The policy is feasible if the instruments of optimization of population are in the hands of the central, not local level. Nevertheless, social welfare in the whole economy is maximal,

$$SW^E = n^E W_A{}^E + (N-n^E)W_B{}^E. \qquad (4')$$

If regional equity is to be respected, the economic instrument can be used. The central government can use in-cash transfers to level up the welfare differences and thus make he solution stable and sustainable. Assume that the transfers are made effectively in a lump sum manner so as to transfer resources and therefore welfare from the richer cities to the poorer ones. This is reflected by a downward shift of the W^A and MW^A curves and an upward shift of the W^B and MW^B curves in Figure 5.

Figure 6 illustrates the effects of lump sum transfers from A-type cities to B-type cities.

(Figure 6 here)

The first and somewhat surprising message of Figure 6 is that the efficient allocation n^E obviously cannot be the target of the transfer policy. This is because the use of a transfer changes the optimality condition. A lump sum transfer from city A to city B leaves the shapes of the *W* and *MW* curves unchanged but shifts the W^A and MW^A curves vertically downwards and the W^B and MW^B curves upwards. As a result the optimality condition given by the intersection point *E* shifts to the left from n^E. An optimal transfer is such that draws the two sets of curves to intersect in point E^* at the allocation n^*, which is the same long-term allocation that is produced by local policy-making. The key difference is that, due to the lump-sum transfer, both the *W* and *MW* curves now intersect at the same population allocation. The solution is both welfare equalizing and efficient. More formally

$$SW^{E^*} = n^* W_A^{E^*} + (N\text{-}n^*)W_B^{E^*} = SW^E. \qquad (4'')$$

Recalling expression (4), it can be concluded that centralized policy intervention by administrative or economic measures not only secures a Pareto efficient outcome, but it also provides a Pareto superior outcome to that yielded by local level policy-making. In terms of total welfare, the effects of the administrative policy measure of quantity rationing and the economic measure of lump sum transfers are equal. In terms of personal welfare and, consequently, in terms of allocation of population, however, the effects are different.

The above finding is very interesting because it contradicts the conventional wisdom of equal effects of the administrative and economic measures. In particular, the result implies that effective transfer policy necessitates variability of the number of cities, precluded that the allocation deviates from the long-term solution n^* yielded by local optimization of population. The latter aspect contradicts the conventional view that inter-regional transfers should, by leveling up the welfare differentials, preserve the status quo of the spatial economy, and thus hinder regional evolution rather than motor it up.

4.3 Further elaboration

The above finding is alarming as to the working of the market mechanism in the spatial context: neither free migration (exit) nor local optimisation of population (voice) seems to secure efficient allocation of population in the economy. Even the long-term result of local policy-making must be altered by centralized transfer policy. The situation is not so bad, however. The result is due to the assumption of heterogeneous cities but more or less homogeneous people. Elaboration of these assumptions gives further support to the working of the spatial market mechanism, at least in principle.

As to the assumption of heterogeneous cities, it might be argued that, in the longer run the initial circumstances in the cities would change and the cities would converge and become more and more homogenous. The possible solutions would then also converge to those presented in Figures 2 and 4. The empirical evidence provided by e.g. the tests of the above mentioned Zipf's law, however, suggests that the basic geographic, climatic etc. differences are quite persistent and hard to overcome, and that the convergence is at the best due to remain partial.

Relaxing the assumption about homogeneity of people and letting them be explicitly heterogeneous in their experiences of welfare yields a more powerful explanation for the existence of market mechanism. The famous Tiebout (1956) hypothesis says that if people are heterogeneous and free to migrate, they can form an effective set of cities, homogenous inside but heterogeneous from each other. As the size of each such city is optimized by maximization of average welfare, social welfare in the economy is maximal. The solution is solely based on the exit and voice type decisions of people, and the role of the central government is totally omitted unless inter-city externalities are allowed to exist.

The Tiebout hypothesis yields a convenient analogy between the Tiebout framework and the familiar competitive market model. Such an analogy is crucial if one wants to translate the message of the abstract competitive models into the language of the real world operating in the spatial context. Due to the extreme importance of the issue, the variants of the Tiebout model are numerous. The existence and efficiency of the Tiebout equilibrium, however, rest on a quite restrictive set of assumptions. From the club theoretic viewpoint the basic preconditions are that the total population is large enough, and that the national territory can support the large number of optimal cities so that all kinds of preferences can be satisfied at the maximum of personal welfare without any integer problems.

The validity of the Tiebout hypothesis in describing the competitive market analogy in the form of an efficient equilibrium distribution of households among jurisdictions (or cities as they are called here) is vividly discussed in the literature. The treatments in the literature are usually more partial that that in this paper, but the main conclusion is that the validity has remained somewhat unresolved both theoretically and empirically. (Stiglitz, 1977; Topham, 1983; Rubinfeld, 1987; Cornes & Sandler, 1986).

Yet, the virtue of the Tiebout hypothesis is in its intuitive appeal. Being aware of the obvious existence problems and the strictness of the assumptions, the argumentation can in principle be regarded reasonable enough to describe the long-term market mechanism in allocating people efficiently in the spatial economy. Unfortunately, the Tiebout case is hard or even impossible to visualise graphically because of the added dimension of heterogeneity of people. The two-dimensional and pair-wise presentation does not suffice to depict city-dependent variability such that is not explained by city size equally for all people. At the simplest, the average welfare curves might be drawn to reach the same peak values for all cities in spite of their intrinsic heterogeneity. The main intuition still is that, at the long-term optimum, the experienced average welfares is equal in each heterogeneous city so that nobody has a motive to emigrate in seek for higher welfare.

5 Conclusions

The club theoretic model of the paper explains the welfare that a city can provide for its residents by city size. Proper utilization of agglomeration economies and diseconomies determines a unique population optimum for each city. Migration based on welfare differentials incorporates the basic market mechanism into the spatial economy, but it cannot self evidently allocate people efficiently in geographic space. Locational choices differ from those in common commodity and factor markets in that people are not able to choose the preferred amounts or visitation rates. This elementary fact somewhat obscures the market mechanism in the spatial context. The local governments must act as supplementary market agents in choosing optimal population. This makes city size a continuous choice variable, and constitutes a close enough market analogy.

The analysis shows that perfect mobility of people will alone yield efficiency only in the special case of homogenous people and big and homogenous cities. But even in this case, a Pareto superior outcome can be reached in the longer run if local governments restrict mobility so as to maximise personal net benefits of their residents. The long-run equilibrium necessitates that the number of cities increases so that optimal utilization of agglomeration economies is reachable in every city. Furthermore, in the case of small cities, migration will lead in corner solutions without local optimisation of population. A Pareto efficient optimum necessitates that the growing cities optimize on their population, and that the falling number of cities is an endogenous variable in the long-term evolution of the spatial economy. Total population can also be taken as a variable, but if total population is allowed to decrease the welfare implications are complicated unless a more global perspective is adopted.

The case of an economy consisting of asymmetric cities is both realistic and interesting. It says that if people are homogenous in that they see the welfare potentials in the cities equivalently and that systematic migration is based on truly experienced welfare differentials a stable and efficient market solution is not at all reachable. In this case even local optimization of population is not sufficient to sustain efficiency in the economy, and centralized intervention is necessary.

The central government has two policy options, administrative regulation and lump sum transfers. The administrative measures, or quantity rationing, can be used to maintain the efficient allocation of people in spite of the existing welfare differentials between more prosperous and les prosperous cities. These kinds of measures may be difficult to implement in a free market economy because they necessitate constraints on mobility.

Transfers between the cities yield a more promising policy measure. Lump sum transfers from richer to poorer cities aim to equalize the welfare differentials. The social welfare effects of these options are the same but, because of the welfare equalizing nature of the economic measure, the effects on individual welfares and on final allocation of people are different. In the transfer option people are allocated so that maximum individual welfare is reached everywhere, given the lump sum transfer. The optimal allocation corresponds to that yielded by the within-club rule of local policy-making, which is to say that less people live in the more prosperous cities than in the regulation option.

The findings about the effects of inter-city transfer policy are interesting because they not only say that the effects differ from those of the administrative policy alternatives but also reveal the fact that effective lump sum transfer policy necessitates endogenous determination of the number of cities. Therefore, lump sum transfers actually motor up reallocation of people between cities, which is a result that contradicts the conventional wisdom that inter-regional transfers should stabilize migration and inter-regional evolution in the economy.

The final remark to the question about the working of the market mechanism in the spatial context is given by the famous Tiebout hypothesis. Assuming people explicitly heterogeneous from each other changes the picture profoundly. If people are allowed to see the connection of city size and welfare differently, and letting them to gather up according to these preferences, then a set of heterogeneous cities with homogeneous populations is formed. If all these cities optimize on their population and no integer problem exists, the outcome is a Pareto efficient long-term solution.

The Tiebout hypothesis is abstract in nature and its practical usefulness is highly controversial, but it still receives serious attention. Its key point is that an efficient market equilibrium is feasible in the spatial context without any intervention by the central government. Of course it must also in this case be precluded that there are no inter-city externalities or other such issues such that would be of interest from the point of view of national-level welfare.

References:

Atkinson, A. & J. Stiglitz (1980): *Lectures on Public Economics.* McGraw-Hill, London.

Bailey, S. (1999): *Local Government Economics. Principles and Practice.* MacMillan Press Ltd, Houndmills.

Buchanan, J. (1965): An Economic Theory of Clubs. *Economica,* vol. 32.

Cornes, R. & T. Sandler (1986): *The Theory of Externalities, Public Goods, and Club Goods.* Cambridge University Press, Cambridge.

Cullis, J. & P. Jones (1998): *Public Finance and Public Choice.* 2nd ed. Oxford University Press Inc., New York.

Fujita, M. & J-F. Thisse (2002): *Economics of Agglomeration. Cities, Industrial Location, and Regional Growth.* Cambridge University Press, Cambridge.

Henderson, J.V. (1985): *Economic Theory and the Cities.* 2nd ed. Academic Press, Orlando.

McCann, P. (2001): *Urban and Regional Economics.* Oxford University Press, Oxford.

Mills, E. & B. Hamilton (1984): *Urban Economics*. Harper Collins College Publishers, New York.

Ng, Y-K. (1973): The Economic Theory of Clubs: Pareto Optimality Conditions. *Economica*, vol. 4, no. 159, August.

Richardson, H. (1973): *The Economics of Urban Size.* Saxon House/Lexington Books, Westmead.

Rubinfeld, D. (1987): *The Economics of the Local Public Sector*. In: Handbook of Public Economics, A. Auerbach and M. Feldstein (eds.). Elsevier Science Publishers B.V. (North Holland)

Scotchmer, S. (1994): *Public Goods and the Invisible Hand*. In: Modern Public Finance, J. Quigley and E. Smolensky (eds.) Harvard University Press, Cambridge Ma.

Stiglitz, J. (1977): *The Theory of Local Public Goods*. In: The Economics of Public Services, M. Feldman and R. Inman (eds.), MacMillan Press, London.

Tiebout, C. (1956): A pure theory of local expenditures. *Journal of Political Economy* 64, October.

Topham, N. (1983): *Local Government Economics.* In: Public Sector Economics, Millward et. al. (eds.). Longman, New York.

Figure 1: Migration between big homogenous cities

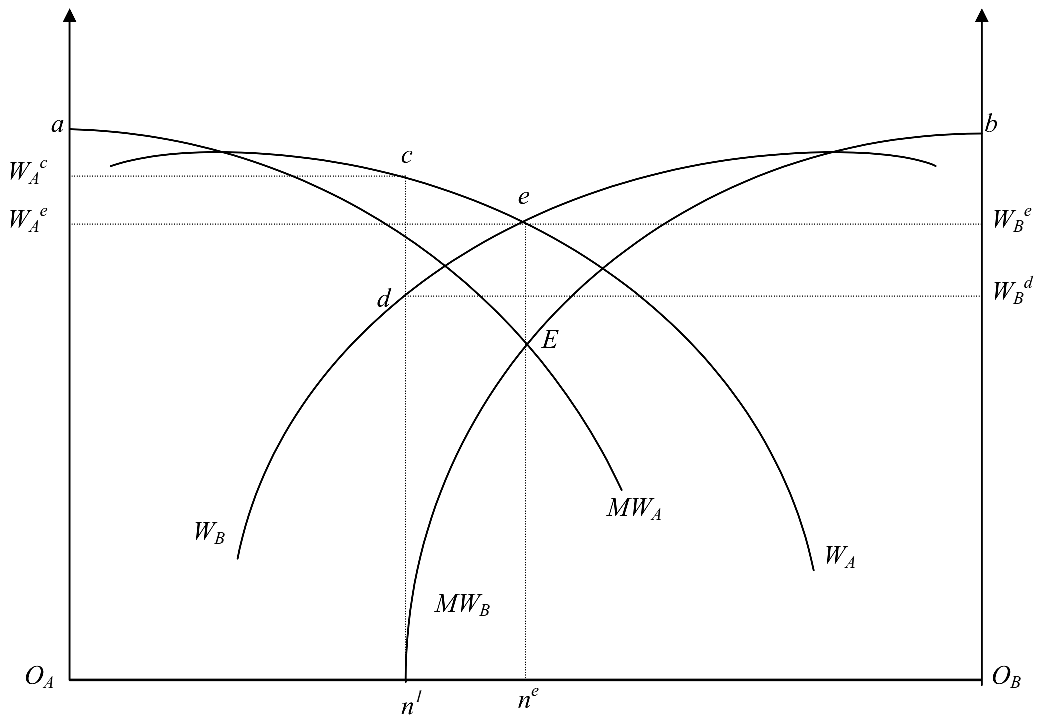

Figure 2: General equilibrium with optimization of city size

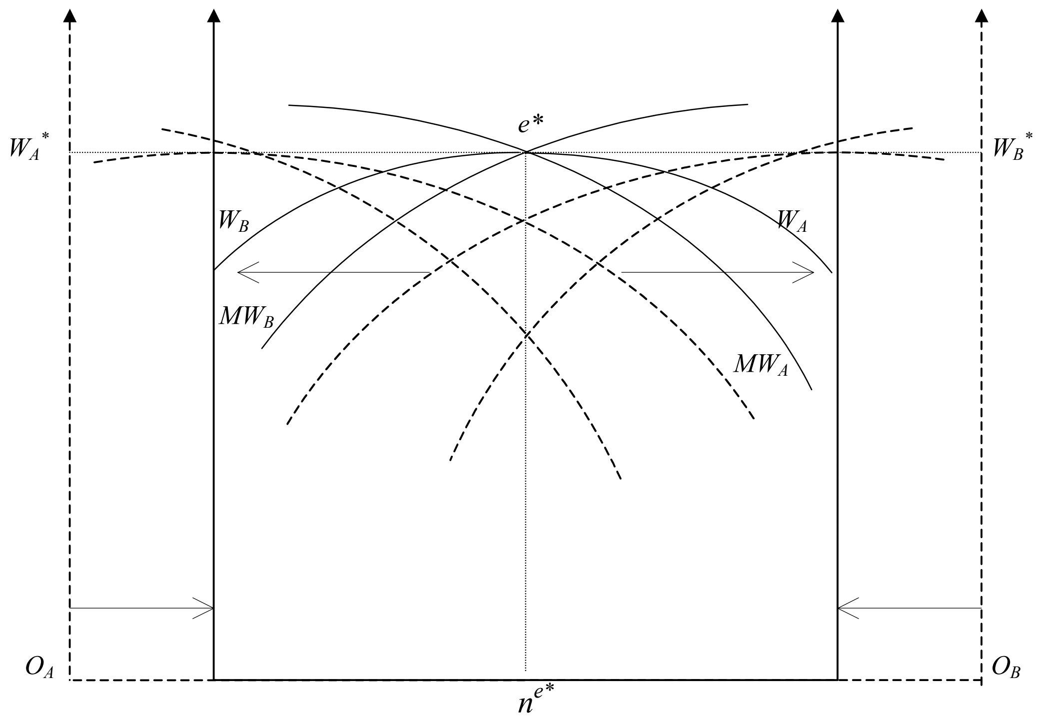

Figure 3: The case of small cities

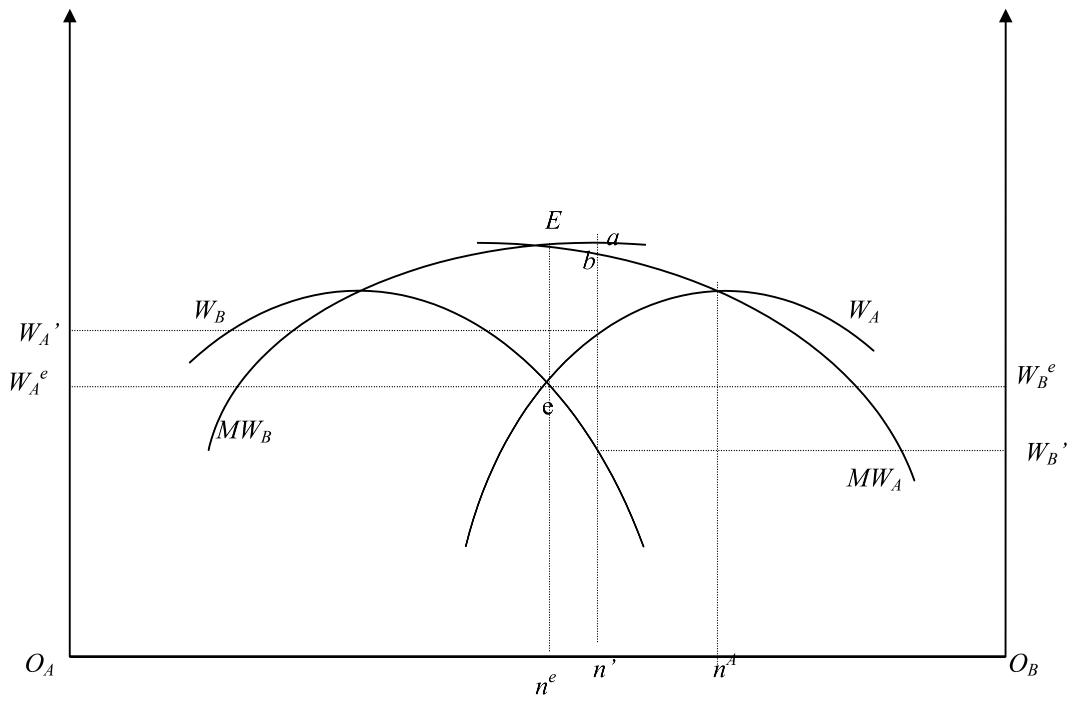

Figure 4: General equilibrium of small cities

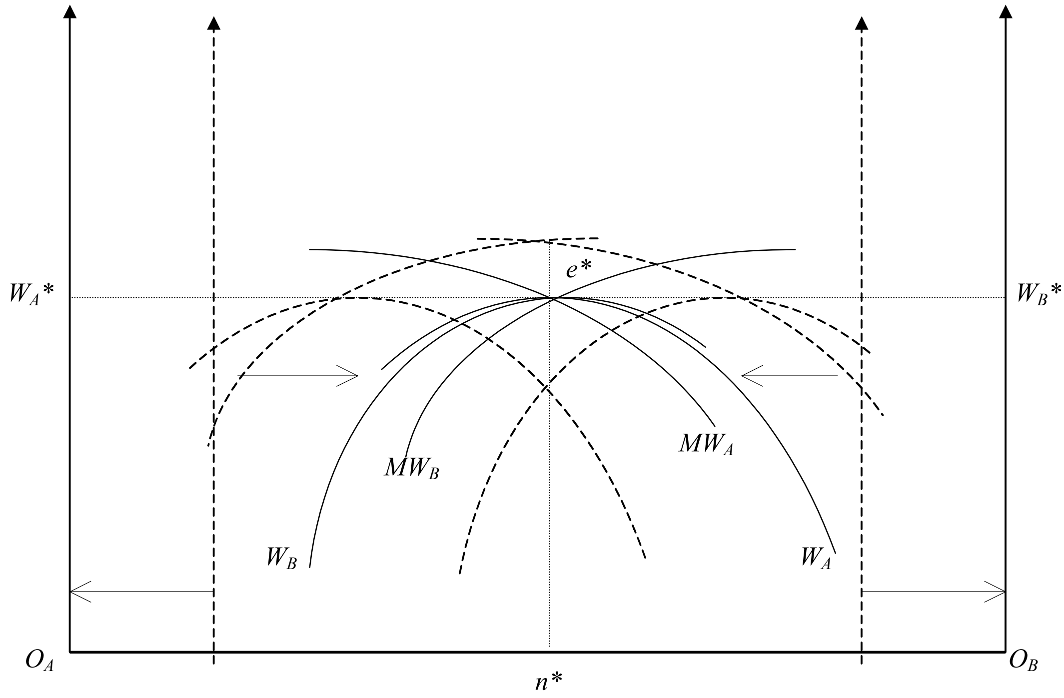

Figure 5: Heterogenous cities

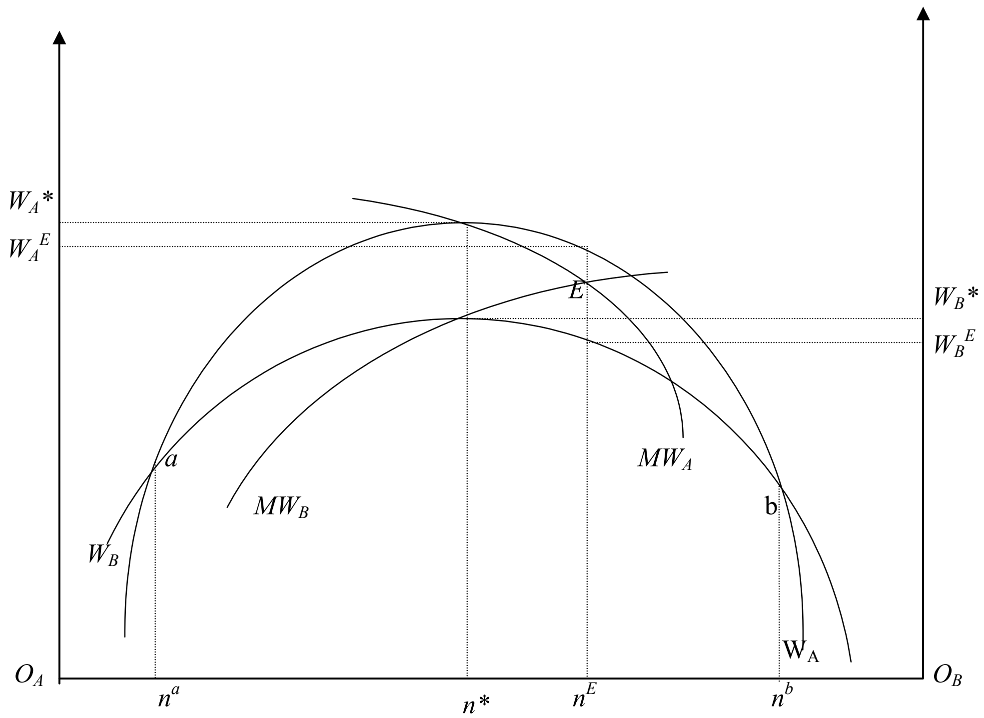

Figure 6: Lump sum transfer from city A to city B