HANNU LAURILA

# Three Approaches to the Economics of Inter-Municipal Migration

■

ACADEMIC DISSERTATION
To be presented, with the permission of
the Faculty of Economics and Administration
of the University of Tampere, for public discussion in the
Paavo Koli Auditorium of the University, Kanslerinrinne 1,
Tampere, on October 27th, 2006, at 12 o'clock.

UNIVERSITY OF TAMPERE

ACADEMIC DISSERTATION
University of Tampere,
Department of Economics and Accounting
Finland

Distribution
Bookshop TAJU
P.O. Box 617
33014 University of Tampere
Finland

Tel. +358 3 3551 6055
Fax +358 3 3551 7685
taju@uta.fi
www.uta.fi/taju
http://granum.uta.fi

Cover design by
Juha Siro

Printed dissertation
Acta Universitatis Tamperensis 1174
ISBN 951-44-6722-1
ISSN 1455-1616

Electronic dissertation
Acta Electronica Universitatis Tamperensis 553
ISBN 951-44-6723-X
ISSN 1456-954X
http://acta.uta.fi

Tampereen Yliopistopaino Oy – Juvenes Print
Tampere 2006

## ABSTRACT

The question of inter-municipal migration is interesting for both empirical and theoretical reasons. Migration has grown more and more extensive in Finland and concerns a wide variety of elementary choices, both individual and collective, which are of immense importance to the functioning of the market economy. This thesis provides theoretical clarification of the issue of spatial resource allocation with special emphasis on the migration of people. The thesis examines the effects of migration into and out of municipalities on the welfares of their current and potential residents and on social welfare. The focus is on economic efficiency, and the main intuition is that migration should be a vital element in the real-world mechanism of resource allocation.

The thesis consists of six essays integrated into a single work, the logical structure of which is given by three approaches called the Partial Equilibrium Approach, the General Equilibrium Approach and the Club Theoretic Approach.

The Partial Equilibrium Approach is based on a synthesis of the classical model of inter-regional labour migration and the model of trade policy. It is found that the short-run effects of labour income taxation on migration and welfare are asymmetric in out- and in in-migration municipalities: taxes imposed in an out-migration municipality cause effects on migration and welfare only if the tax is decidedly high, but taxation in an in-migration municipality causes out-migration and causes welfare losses even for a small tax. However, the effects remain minor if the out-migration municipality is small and rich in capital, and/or if the in-migration municipality is big and rich in capital. Thus, at least in Finland, the short-run effects remain quite insignificant in practice.

The General Equilibrium Approach applies the seminal neo-Keynesian macroeconomic model to the analysis of inter-regional factor migration. Four types of market adjustment are specified, namely price adjustment, nominal wage adjustment, opposite adjustment and parallel adjustment. The effects of labour migration in particular on local interest rates are shown to depend on the specific adjustment path. From a theoretical point of view, an interior solution of spatial resource allocation is not in general granted. The model is used to examine the effects of local policy on migration, and an AEH-type explanation is used to describe people's perceptions of the policy. A fair policy programme, which does not distort the real wage comparisons between localities, does not affect migration. On the other hand, an unfair programme is shown to encourage migration the more the farther the anticipations concerning the repayments are from the true nature of the program. In the longer term, when people correct their perceptions concerning the repayments, the effects are mitigated and eventually reduced to zero, provided that the programme is actually fair.

In the Club Theoretic Approach, welfare in a city depends on the size of the city because of agglomeration economies and diseconomies. Technological and pecuniary externalities are internalised in the city and explain migration and the formation of cities. It is shown that migration may lead to non-optimal solutions, even when the city itself acts as a true market agent in choosing its optimal size from a within-club viewpoint. The results of the effects of national level policy somewhat contradict the conventional wisdom. First, administrative and economic policy measures are shown to differ in their effects, and second, if cities are initially outside their within-club optima, lump sum transfers rather accelerate than stabilise spatial evolution. The competitive dynamics of the system is a product of people's exit-type choices (migration) and their collective voice-type choices. The use of local efficiency enhancing measures depends on the pressure of exit. A necessary condition for dynamic efficiency is that the migration equilibrium is non-stable. Dynamic efficiency is more likely to emerge between initially small than initially large cities. The pressure for efficiency is weaker in the most relevant case of asymmetric cities.

## TIIVISTELMÄ

Kuntienvälinen muuttoliike on mielenkiintoinen tutkimuskohde sekä empiirisistä että teoreettisista syistä. Muuttoliike on Suomessa viime aikoina voimistunut, ja siihen liittyvillä yksityisillä ja kollektiivisilla päätöksillä on keskeinen merkitys talouden toiminnan kannalta. Tässä työssä pyritään luomaan teoreettisesti selkeä kuva talouden voimavarojen alueellisen allokoitumisen periaatteista erityisesti kansalaisten muuttopäätösten näkökulmasta. Työssä tarkastellaan muuttoliikkeen syitä ja vaikutuksia itsensä muuttajien, lähtö- ja tulopäiden muiden asukkaiden ja koko yhteiskunnan hyvinvoinnin näkökulmasta. Lähtökohtana on, että muuttoliike on tärkeä osa reaalitalouden voimavarojen kohdentumisen mekanismia.

Työ koostuu kuudesta erillisestä esseestä, jotka muodostavat lähestymistapojensa puitteissa johdonmukaisesti etenevän kokonaisuuden. Lähestymistavat ovat Osittaistasapainolähestymistapa, Kokonaistasapainolähestymistapa ja Klubiteoreettinen lähestymistapa.

Osittaistasapainolähestymistapa perustuu muuttoliikkeen klassisen työmarkkinamallin ja kansainvälisen talouden kauppapolitiikkamallin synteesiin, jonka avulla on aiempia malleja tarkemmin mahdollista tutkia politiikan muutto-ja hyvinvointivaikutuksia. Tärkeä havainto on, että työtuloveron lyhyen tähtäyksen vaikutukset ovat epäsymmetriset: muuttotappiokunnassa veroilla on vaikutuksia vain kun vero on huomattavan korkea, kun taas muuttovoittokunnassa pienikin vero saa aikaan ulosmuuttoa ja hyvinvointitappioita. Vaikutukset jäävät kuitenkin pieniksi niissäkin korkean verotuksen muuttotappiokunnissa, jotka ovat pieniä ja pääomarikkaita, sekä niissä muuttovoittokunnissa, jotka ovat suuria ja pääomarikkaita. Ainakin Suomessa verotuksen lyhyen tähtäyksen vaikutukset ovat siis käytännössä vähäiset, toisin kuin yleensä tunnutaan ajateltavan.

Kokonaistasapainolähestymistavassa sovelletaan Neokeynesiläistä makromallia tuotantopanosten alueellisen liikkumisen kokonaistaloudellisten vaikutusten tutkimiseen. Mallin sopeutumisurat tyypitellään kolmeen luokkaan: hintasopeutus, nimellispalkkasopeutus, vastakkaissopeutus ja samansuuntainen sopeutus. Sopetumistavalla on merkitystä erityisesti korkotason kannalta – korot voivat nousta tai laskea riippuen talouden kysyntäpuolen olosuhteista, jotka määrittävät sopeutumisen kulun. Sisäratkaisun löytyminen voimavarojen aluelliselle kohdentumiselle ei ole mallin mukaan taattua, vaan kulmaratkaisuja saattaa esiintyä. Paikallisen politiikan vaikutukset muuttamiseen ja hyvinvointiin riippuvat sitä koskevista odotuksista. Politiikka, jossa perityt verot palautetaan täysimääräisinä, ei vaikuta reaalipalkkaan eikä muuttamiseen, mutta mikäli näin ei ole tai jos ihmiset arvioivat väärin saamansa vastineen, muuttokannustimia esiintyy. Muuttokannustimet lievenevät ja poistuvat kokonaan pidemmällä tähtäyksellä, kun ihmiset vähitellen korjaavat käsityksiään politiikan suhteen.

Klubiteoreettisen lähestymistavan mukaan kaupungin kyky luoda hyvinvointia asukkailleen riippuu kasautumisen etujen ja haittojejn vuoksi sen koosta. Tekniset ja pekuniaariset ulkoisvaikutukset sisäistyvät muuttopäätöksissä ja ohjaavat kaupunkien kehitystä. Muuttoliike ei välttämättä yksin tuota tehokasta lopputulosta, ei edes silloin kun asukkaat kollektiivisesti päättävät kaupungin optimaalisesta koosta. Tärkeä havinto on, että kansallisen kaupunkipolitiikan hallinnolliset ja taloudelliset ohjauskeinot ovat vaikutuksiltaan erilaiset ja valinta niiden välillä on tehtävä politiikan päätavoitteen mukaan. Kiinteäsummainen tulonsiirto ei ole muuttoliikkeen kannalta neutraali, kuten yleensä on totuttu ajattelemaan, vaan se pikemminkin aiheuttaa kuin stabiloi muuttamista. Kilpailulliset olosuhteet kaupunkien muodostamassa kokonaisuudessa syntyvät paikallisen itsehallinnon ja muuttoliikkeen keskinäisyhteydestä. Muuttoliikkeen uhan alla etenkin pienet kaupungit joutuvat kilpailun paineessa toimivien yritysten tapaan jatkuvasti tekemään asukkaidensa hyvinvointia parantavia toimia, mikä takaa dynaamisen tehokkuuden toteutumisen taloudessa.

CONTENTS

# 1. INTRODUCTION

## The research object

The question about inter-municipal migration is interesting for both empirical and theoretical reasons. Empirically the question has become more and more important in Finland in recent years because of the major structural change that the economy has been going through. Regional development is now based on the market mechanism much more directly than before. Economic motives for migration have reinforced because the mobility of jobs has increased and regional disparities have widened in the new economic circumstances.

At the same time, national policy concerning regions and municipalities has undergone change, and the financial focus of local governance has been shifted from state dependency towards locally determined tax bases. The fact that the elements of welfare are nowadays clearly in urban surroundings has turned the Finnish migration pattern from rural-urban to inter-city migration. The changes urge the need for economic research on the mechanisms of market-driven spatial evolution and on the role and instruments of local policy-making.

The issue of migration is particularly interesting from the point of view of economic theory. First, the classical economic theory usually treats the concept of individual welfare, or utility, very abstractly. In the real world individual utility is naturally very place-dependent. Utility is experienced in everyday life, in a spatially determined sphere of consumption transactions and other activities. On the other hand, the incomes that facilitate the generation of consumption utility originate on the factor markets, especially on the labour market. The labour market is obviously very local in nature even in the present stage of economic development. Since local commodity and factor markets are the main bases of utility, utility maximisation must include seeking for the best combinations of these interdependent marketplaces.

Second, the classical theory is very abstract on space-induced transaction costs in the marketplace, and the classical models also ignore economies of scale, externalities, imperfect competition and other issues such that explain the existence of urban agglomerations. In the real world, however, it is quite undeniable that urban agglomerations are at the heart of a working market mechanism. The relationship between urbanisation and economic development is clear. The newer literature in urban economics has provided valuable insights not only on the formation and functioning of the cities themselves but also on the value of cities and their networks to the whole economy.

Third, one of the key issues of the classical theory is the efficiency of resource allocation. The concept of resource allocation is again treated abstractly in the classical models. Yet the concept is very concrete in the real world, where firms continuously seek for profit maximising locations, and where consumers migrate in order to maximise their welfare. Almost any reallocation of resources must imply physical movement of production factors. Therefore, factor migration is of immense importance to economic efficiency, and the functioning of the real world allocation mechanism is in a focal position when the classical theory is applied in practice.

Fourth, the most abstract versions of the classical theory ignore the existence of public goods. Again, public goods, pure and less pure, do exist in the real world, and a vast majority of them is also local in nature. Taking this into account, economies of scale and scope, transaction costs, externalities etc. state that the provision of local public goods should be carried out in more or less densely populated surroundings, where people derive their utility from private and public consumption and other everyday pastimes.

Fifth, in their localised life people not only choose between local public economies by migration but they also collectively steer the preconditions of their welfare in their present locations. At least in Finland local jurisdictions, which derive their power directly from their residents, are autonomous enough to decisively affect the elements of locally experienced welfare. This can be done not only through designing attractive tax-service packages but also through amending the conditions of the local private markets. The jurisdictions are actually able to compete for residents and jobs and they can be analysed as true market agents composed of welfare maximising individuals.

To sum up, analysis of inter-municipal migration actually turns out to be more than just explaining why people migrate – it concerns a wide variety of elementary choices, both individual and collective. The choices are quite complex and they are of immense importance to the functioning of the market economy. The research object thus serves an inexhaustible source of interesting questions for economic research, of which only a small fraction is studied in this thesis. Thus, the dissertation is not a comprehensive survey of all the relevant issues, rather the opposite.

### Aims and outlines of the thesis

The thesis aims to provide theoretical clarification to the complex issue of factor migration with special emphasis on migration of people. The thesis makes use of advanced economic theory to examine the effect of migration into and out of municipalities on the welfares of their current and potential residents and on social welfare. The thesis concentrates on how utility is affected by the location of work and residence through the locational characteristics of leisure, the prices and availability of private and public goods, real wage rates etc. The focus is on migration from the point of view of efficiency, and the main intuition behind the analysis is that migration should be an important element in the real-world mechanism that leads to a technically and allocatively efficient market equilibrium.

The main goal of the thesis is to provide original insight into migration of people and to set up theoretical textbook-type general frameworks that can be utilised in a detailed analysis of policy effects and other experiments. Some examples of such applications are provided to show that new theoretical breakthroughs can be expected. The thesis also paves the way for collecting empirical data and documentary evidence, and using econometric models to empirically identify and measure the issues that are raised.

The thesis consists of six essays originally written as independent texts. They are, however, integrated into a single thesis with a logical proceeding provided by the three approaches presented below. Unnecessary repetition is avoided. Throughout the thesis graphical argumentation is used as the main device of analysis and mathematical argumentation is kept to a minimum.

One of the main assumptions made in the thesis is that the firms in the production sector act competitively and make rapid decisions on production and location so as to maximise profits. Imperfect competition is thus omitted, but economies of scale and scope as well as externalities may exist in production. The firms' optimal decisions are taken to be reflected to people in terms of market information so that all kinds of indirect agglomeration economies are pecuniary in nature. Thus, the analysis of firms' locational choices and the consequent effects on formation of municipalities and on welfare can be omitted.

Second, although it is recognised that housing is a very important element in peoples' lives both in its direct effect on utility experiences and indirect effects through living costs, closer analysis of the housing market is not included. Likewise, location and migration within municipalities, that is the formation of the internal structure of the municipalities is not considered.

Third, it is taken that the municipalities are free from historical, political, administrative or other constraints so that they can form on purely autonomic grounds. In the spatial context, efficiently functioning local private markets together with efficient provision of local public goods constitute the concept of local public economy, which is the economically relevant definition of a financially and administratively autonomous municipality.

Fourth, it is assumed that an ideal model of fiscal federalism is applied in the assignment of national and local public functions. Fiscal federalism is a crucial element of efficiency from the national viewpoint, which is well documented in the literature of local public economics. In this thesis, closer analysis of fiscal federalism is ignored and neither inter-locality externalities nor equity considerations are taken into account.

Fifth, it is simply assumed that the collective decisions are effective both on the local and on the national (or even sub-national) level. As is well articulated in the literature on public choice, there is good reason to doubt this assumption. The complications caused by inefficient public choice should not, however, impair the main messages of the present thesis, at least not more than the fact that the real world is always a second-best, not a first-best.

### The migration of people

The theoretical construction of the thesis rests on describing and explaining migration of the labour factor including human capital and other such factors that are embodied in people. Concentrating on migration is justified by assuming that the firms efficiently optimise on their locations and thus give impulses to people's choices in the form of market price information. People migrate between local

public economies, which are based on private and social economies of agglomeration. The relevant economic areas are thus more or less explicitly attached to cities and urban areas.

The thesis elaborates standard textbook models in order to yield illustrative comprehension on the rather complex set of causes and effects of inter-municipal (and international) migration of people. The thesis provides three elementary approaches to the matter, each of which having their own strengths depending on the particular questions to be addressed. All three approaches are based on the basic model of consumer's choice. People are assumed to choose their places of residence in order to maximise a problem of constrained utility maximisation of the following type:

(1) $Max\ U(q_x,q_y,l)\ \ s.t.\ \ w(1-l) + k = p_x q_x + p_y q_y$.

In expression (1), individual utility is assumed to be positively affected by the arguments $q_x$, $q_y$ and $l$. Argument $q_x$ refers to the quantity of consumption of purely private goods and services, and $q_y$ refers to that of public goods and services, assumed to be (mainly) locally provided. The argument $l$ refers to leisure time. These are the endogenous variables. The standard properties concerning the shape of the utility function are assumed.

In the budget constraint of expression (1), $w$ is the market rent (or wage) for labour time $1-l$ (total available time being normalised to unity), $k$ is non-labour income, and $p_x$ and $p_y$ are respectively the market prices for private and public goods. The prices, including the nominal wage, are the exogenous market parameters. The budget constraint states that all income is spent on private and collective consumption during the time period considered.

Employment is in the core of people's migration choices for several reasons. First, employment opportunities, the search for better jobs, education and cultivation of special skills, building up careers and other such work related issues constitute a major part of the determinants of average people's location decisions in practice. Second, the connection between work and residence is beyond question and, at least from the economic point of view, the concepts of labour market area and residential area match accurately enough. To generalise even further, the concept of local public economy can be applied. And third, the setting is convenient because the real wage, or the purchasing power of income, is a simple approximation of individual welfare. This is to say that the migrants' utility functions can be even omitted in the analysis, and only the budget constraints that determine consumption utility can be considered.

To see the relevance of the classical labour market model, assume that both private and public consumption are measured by $q$ and that their prices are measured by $p$. Ignoring non-labour income, expression (1) reduces to

(2) $Max\ U(q)\ \ s.t.\ (1\text{-}l)\ w = pq$.

Solving $q$ from the budget constraint and substituting yields $U(q)= U[(1\text{-}l)w/p]$. The individual's utility maximisation problem thus reduces to a problem of optimal time use. In its simplest form, when labour time is assumed to be fixed to unity the result is that individual welfare is given (and measured) by the exogenously determined real wage. In any case, it suffices to consider the market conditions given by the budget constraint. The analysis of migration can then be made operational by assuming that real wages are locally determined, and that there may be local differences in the market parameters $w$ and/or $p$. The simple version of the model not only facilitates a compact

treatment of a complicated issue but it also uses concepts that can be empirically observed and measured.

From the perspective of the classical labour market model, problem (1) can be used as a basic element in migration analysis with quite straightforward specifications. There is good reason to accept, however, that the endogenous variables in problem (1) are also place dependent. This complicates the matter somewhat and makes the problem quite unsolvable without further elaborations. This thesis provides three approaches to migration starting from the simplest labour market version and ending with a more comprehensive one, where both exogenous and endogenous variables vary between locations.

## The three approaches

The three approaches to migration (1) are called Approach I: *The Partial Equilibrium Approach* (including Essay #1), Approach II: *The General Equilibrium Approach* (including Essays #2 and #3) and Approach III: *The Club Theoretic Approach* (including Essays #4, #5 and #6).

**Approach I:** ***The Partial Equilibrium Approach*** elaborates the standard migration analysis based on the classical labour market model of migration. Essay #1: *Short-Run Effects of Taxation on Inter-Municipal Migration* constructs a merger of the labour market model and a model of international trade policy. The classical treatment of aggregated labour markets states that the budget constraint in (2) simply converts to

(3) $Lw = pq$,

where $L$ denotes aggregate labour measured in terms of labour time units of the whole labour force. Regarding migration, labour time units are assumed to be directly convertible to population numbers. The analysis is based on comparisons of labour market conditions and the consequent real wage differences between municipalities, to which people react by migration. In the essay, the model is used to investigate the short-term effects of local labour income taxation on migration and welfare.

**Approach II:** ***The General Equilibrium Approach*** widens the perspective of migration analysis to include also goods and capital markets. Essay #2: *The Macroeconomic Model of Migration* constructs a neo-Keynesian macroeconomic model of migration. The supply side of the model is based on the simple aggregation of the labour market given by Equation (3) above together with a standard neoclassical production function. The demand side is given by the Keynesian IS-LM model. Migration has an obvious effect on the supply side, but Equation (3) implies that aggregate demand depends on migration, too. This is because both $L$, the number of consumers in the local marketplace and $w/p$, the consumers' purchasing power are exogenously affected by migration. Migration is again based on local differences in the market parameters, but the adjustment process is described according to the full general equilibrium. The analysis is completed by allowing capital also to be mobile and investigating the interplay of the two simultaneously mobile factors. Moreover, Essay #3: *Taxes, Transfers and Migration* applies the model to investigate the effects of local policy on migration. A specific tax-transfer programme is incorporated into the model, but the policy construction can be read more generally to concern the share of the public sector in the local economy. The Adaptive Expectations Hypothesis is applied to explain short-term and long-term effects of the policy.

**Approach III:** ***The Club Theoretic Approach*** allows both exogenous parameters and endogenous variables in (1) to be place-dependent and affected by externalities caused by agglomeration.

Therefore, they can be expressed as functions of city size, measured by population. In the above two approaches, local agglomeration economies are assumed to be pecuniary-type externalities, which are internalised in local prices and nominal wages and are thus reflected in individual decisions on residential location. Here technological externalities that enter directly into the utility function are also taken into account. Essay #4: *The City as a Club* constructs the club theoretic framework. The conversion to the club theoretic model is done by writing $q_x = q_x(n)$, $q_y = q_y(n)$, $l = l(n)$, $w = w(n)$, $k = k(n)$, $p_x = p_x(n)$ and $p_y = p_y(n)$ in expression (1), with $n$ denoting city size. The sets of private and collective goods thus depend on city size as does the amount and quality of leisure time. In the budget constraint, city size affects wages and all consumption prices. The non-labour incomes may be assumed to be size-dependent as far as the required market transactions are spatial in nature. The club-theoretic property of the city states that agglomeration economies and diseconomies are transformed into locally experienced welfare. Since all the endogenous arguments in the utility function and all the variables and parameters in the budget constraint depend on $n$, problem (1) can be rewritten in a club theoretic form as

(4) $Max\ W\ s.t.\ W = B(n) - C(n)$,

where $W$ is individual welfare (or net benefit), $B$ is the individual benefits experienced in the city and $C$ is the individual net monetary costs of living in the city. The benefit side $B$ refers to the utility function and the cost side $C$ refers to the budget constraint in expression (1). Technological externalities affect the benefit side, and pecuniary externalities affect the cost side. A conventional assumption is that positive externalities (or agglomeration economies) dominate in the earlier stages of city growth, but that negative externalities (or agglomeration diseconomies) start to dominate in overcrowded surroundings. Thus there exists a theoretical optimal size for the city.

The approach also incorporates collective choices *(voice)* into individual decision-making along with the decisions on residential choice *(exit)*. Essay #5: *Migration, Policy and the Formation of Cities* makes use of the model of Essay #4 to show the conditions of efficient formation of cities and spatial resource allocation in an economy consisting of a multiple of cities. The incorporation of *voice* is valuable for two reasons. First, it is the collective, not an individual migrant, who makes the market-type choice on quantity (that is population) in the model. Second, by collective choice, the residents of a locality are able to directly affect their welfare by altering the conditions of local public good provision and by improving the functioning of the local private market. Taking this perspective, it is possible to see the match between the concepts of local market area and local public economy in the spirit of the seminal Tiebout model of efficient provision of public goods. Essay #6: *Urban Governance, Competition and Welfare* extends the analysis even further to demonstrate the competitive pressures behind local urban governance. In the club theoretic model the local public economies are, under certain circumstances, subjected to competition like private firms so that sufficient conditions for technical efficiency are granted.

# 2. APPROACH I: THE PARTIAL EQUILIBRIUM APPROACH

## Essay #1: Short-Run Effects of Taxation on Inter-Municipal Migration and Welfare

***Abstract:*** *Short-run effects of taxation on inter-municipal migration and welfare are examined in a competitive labour market model. The model is constructed as a synthesis of a classical model of inter-regional labour migration and a model of trade policy. It is found that a labour income tax in an out-migration municipality causes effects on migration and welfare only if the tax is decidedly high. Furthermore, if the municipality is small and capital intensive in its industrial structure, the effects are slight even for a high tax. Taxation in an in-migration municipality, however, reverses the migration pattern and causes welfare losses even for a small tax, but the effects are slight if the municipality is large and capital intensive. There is an asymmetry in the effects of a given tax change. Nevertheless, the effects remain fairly insignificant in practice, at least in Finland.*

*Key words: classical labour market model, migration threshold, tax competition*
*JEL classification: 931*
*An earlier version of the paper was published under the title 'Verotus ja kuntien välinen muuttoliike' in: Finnish Local Government Studies 3/2002.*

### 1 Introduction

The Finnish economy has undergone a major structural change in recent years. One feature of the change is that regional development is now much more directly based on the market mechanism. Economic globalisation has emphasised the role of absolute cost advantages as a determinant of industrial location, which in turn has resulted in increasing mobility of jobs. Even in its present stage Finnish structural change is largely about a transition from a rural-based to an urban-based economy.

The change in the economic environment has also caused considerable changes in the institutional preconditions of regional and municipal development. EU integration has changed the tenets of regional policy, and the financial focus of local governments has shifted from state grants towards their own, locally determined tax bases. Inter-municipal migration has been increased after a long period of stability, and municipalities have been divided into those of in-migration and out-migration.

The revival of migration has provoked a lively discussion. One line of the discussion has concentrated on the effects on migration of taxation. A prominent reference has been the 'conventional wisdom' of the economics of migration, namely that labour income taxation distorts economic decision-making on residential location, and induces people to avoid taxes by voting with their feet. An ultimate interpretation of this argument is that inter-municipal and even international differences in tax levels are not sustainable, and municipalities and countries will be forced to equalise their tax rates.

Some attention has also been paid to the effect of migration on local public finance. From a long-term perspective it is evident that, in the new institutional environment, emigration erodes the local tax base and exerts pressure to raise the local tax rates, at least if certain standards in public services must still be maintained. On the other hand it is also evident that immigration strengthens the local tax base and eases the pressure for tax increases. But it has also been pointed out that in the shorter term migration may cause costs and create tax pressures in both in- and out-migration municipalities.

In any case, there seems to be a common understanding that taxation is an important factor in migration and in the division of municipalities into those with in- and out-migration. Out-migration municipalities are easily regarded as doomed to a vicious circle in which emigration causes tax increases, which in turn speed up emigration, which causes further pressure for tax increases, and so on. Moreover, migration is often thought to be a zero-sum game. According to this view, concentration in the biggest growth centres seems not only inevitable, but also tax-driven.

The question about the effects of taxation on migration is interesting not only because of its practical importance but also from the point of view of economic theory. In the public discussion the economic intuition that taxes inevitably accelerate emigration seems to be widely accepted.

However, in order to be on solid ground when using this ‘conventional wisdom’ as an argument in the tax debate, the basic message of economic theory must be carefully reviewed.

This paper describes the short-term effects of taxation on inter-municipal migration and welfare. For the analysis, a simple inter-municipal labour market model is constructed, in which the decisions of migration are based on real wage differences between municipalities. The municipalities are assumed to be fiscally autonomous and to use labour income tax revenue for their fiscal purposes. The partial equilibrium model presented is a novel synthesis of a labour market migration model familiar from regional economics, and a model of trade policy familiar from international economics. The synthesis provides a fuller understanding of the impact of inter-municipal differences in tax rates.

Because of the basic nature of the model, the main conclusions derived from it are most general and therefore reflect economic intuition that can most reasonably be called the ‘conventional wisdom’ of economics. The analysis aims to shed light both on the question of the effects of taxation on migration and welfare and on the relevance of the commonly used argument of tax-accelerated migration.

# 2 The model

## 2.1 Elements of migration

In a market economy people choose those residential locations that are optimal for their personal purposes. The choice may be seen as a solution to a conventional utility maximisation problem. People’s utility and so their residential choices are affected by a wide variety of economic and non-economic factors, most of which are more or less place-dependent (Laurila, 2000). In order to simplify the setting it is reasonable to focus on the most relevant of these place-dependent factors.

The simplest way to model residential choices is to concentrate on purely economic factors. The classic theory of labour migration is constructed on labour market analysis (Armstrong & Taylor, 2002, pp. 141-143; McCann, 2001, pp. 175-207). This is justified by both empirical and theoretical reasons. Income from labour is the most important base of consumption options for average people, and the purchasing power it provides (or the real wage) is a simple and commonly used measure of individual welfare. The spatial relevance of the labour market model is obvious, because real wages are determined on locally characterised labour markets.

The key feature of the classic theory of labour migration is that all the variation that determines people's residential choices is treated in terms of the exogenous market parameters. That is, as all the variation enters the budget constraint, the utility function can be omitted. On the income side of the budget constraint, labour income depends on nominal wages and on labour time. Labour time is an endogenous choice variable, but the nominal wage rate is exogenous in the short term, and may vary between municipalities according to local conditions. The expenditure side of the budget constraint is given by consumption choices determined by the market prices of private and public goods and services. These price parameters may again differ between municipalities depending on local market conditions.

To formalise, ignore non-labour income, normalise total available time to unity, compress private and public goods and services into one consumption bundle, and write the following individual maximisation problem:

(1) $Max\ U(q,l)\ s.t.\ (1-l)w = pq,$

where $q$ is the amount of private and public goods and services consumed, $l$ is leisure time and $(1-l)$ is labour time, $w$ is the exogenous nominal wage, and $p$ is the exogenous price of the consumption

bundle. Assume that the qualitative aspects of both leisure time and the consumption bundle are fully reflected in the market price information, and assume that the prices are locality-dependent. Then, as the price parameters *w* and *p* determine the income side and the expenditure side of the budget constraint, they also determine the welfare level that can be derived from consumption. Solving consumption from the budget constraint yields

(2) $q = (1-l)\omega$,

where $\omega = w/p$ is the local real wage, that is the local nominal wage *w* deflated by the local price level *p*. Substituting (2) back into the utility function in (1) shows that the problem reduces to a mere question of optimal use of time. In this simple framework, it is the possible variation in the exogenous real wages between localities alone that gives reason for inter-municipal migration. People simply migrate from low-real-wage municipalities to high-real-wage municipalities.

### 2.2 The labour market

Local real wages are determined on competitive local labour markets. The production technology is given by the function

(3) $q^i = f^i(K^i,L^i)$,

where *K* denotes capital and *L*, composed as a sum of individual efforts, denotes labour time. The superscripts refer to the locality considerd, and they imply that the production factors are rival between localities, and that the technology may differ between municipalities because of existing industrial structures. The capital stock is assumed to be constant in the short term so that labour is the only variable factor. The usual neoclassical assumptions on the production function are made, including diminishing marginal productivity of labour. People and firms are assumed to have equal

information about the fundaments of the economy. That is to say that the real wage determines both demand and supply on the labour market. (Burda & Wyplosz, 1997, s. 136-150.)

Assume that municipalities are formed on an economic basis so that they geographically cover economically relevant labour market areas. Furthermore, working in one municipality necessitates residence in that municipality. The firms are assumed to be tied to their locations in the short term, and the constant capital stock is assumed to be immobile between the municipalities. Therefore, labour is the only mobile production factor in the economy. Labour earnings and possible public transfers are determined locally, but capital incomes are assumed to be locally independent.

The price level is determined on a local basis. Because not all goods and particularly not all services are tradable between municipalities, local production is mostly sold to the local market. Housing costs are clearly very local in nature because real estate, buildings and technical infrastructure are totally immobile in the short term. The costs of local public services, namely taxes and tariffs, constitute an important part of determination of the local price level. These costs are purely local in nature, because public services are mostly non-tradable and because there are no inter-municipal transfers or state grants by assumption.

Figure 1 illustrates the standard neoclassical labour market model. The figure includes demand and supply of labour time and the competitive market equilibrium on the local labour market.

(Figure 1 here)

In Figure 1, the real wage is measured on the vertical axis, while the horizontal axis measures labour input in terms of time units. Under the assumption of constant capital stock, and by the properties of the production function, labour demand is determined by a falling schedule $D_L = MPP_L$, where

$MPP_L$ is the diminishing marginal physical product of labour. The labour demand curve is a horizontal sum of the demands of the local firms at each real wage level, based on their industry-dependent abilities to recruit labour. An exogenous increase in the capital stock would shift the demand curve outwards and vice versa. The slope of the demand curve reflects the wage sensitivity of labour demand, which again depends on industrial structure. The more capital intensive the local industrial structure is, the steeper is the labour demand curve and vice versa.

The supply of labour is determined by people's time use decisions as individual solutions to problem (1). In Figure 1 the schedule $S_L = MC_L$, where $MC_L$ is the opportunity cost of leisure (or the marginal cost of time), presents the local supply of labour. The schedule is a horizontal sum of individuals' time use decisions for every real wage level. It is assumed that the market equilibrium occurs on the rising regime of the local supply curve, which is to say that a rise in the real wage both motivates people to work longer hours and attracts new people to the labour force and vice versa. The possibility that the supply curve may bend backwards when the income effects start to dominate in people's optimal time use decisions is ignored.

On the competitive labour market the real wage adjusts perfectly either upwards or downwards so that labour demand meets labour supply in a full employment equilibrium. In Figure 1 the equilibrium is at the intersection point, point *e*, obtained up by the adjustment of the real wage to $\omega^e$. Employment, measured in time units, is $L^e$. The segment of $S_L$ above $\omega^e$ indicates labour time that remains unused in production because people find it optimal to spend it on leisure, given the real wage $\omega^e$. Under the assumption of continuity in labour supply this segment is a sum of individual time use decisions, including those who remain totally outside the labour force, those who work part-time, those who refuse to work overtime etc. The segment of $D_L$ below $\omega^e$, in turn, indicates production possibilities that do not materialise because the physical product of labour is, at the present capital stock, too low compared to the existing real wage.

The labour market equilibrium is stable because the adjustment stops both from below and from above at the real wage level $\omega^e$ in Figure 1. The equilibrium is also efficient, because social welfare, measured as the product of labour by area *abe* between labour demand and supply curves, is maximal. Of this area, the upper part $w^e be$ is the firms' surplus from securing employees up to point *e* cheaper than their marginal physical product, and the lower part $aw^e e$ is the workers' surplus from earning higher real wages up to point *e* than what their reservation wage would have been. Any other real wage than $\omega^e$ would make the areas and so social welfare smaller.

### 2.3 Inter-municipal migration

Given that real wages are determined in local labour markets and allowing for a set of municipality-level local markets, the labour market of the whole economy consists of the migration-driven integration of the local markets. To analyse inter-municipal migration, assume that in choosing their optimal residential locations, people are apt to monitor possible inter-municipal differences in real wages, and suppose that there are no transaction costs or other frictions that would distort the decision-making. Since working in another municipality is not possible without moving, migration is triggered by possible differences in local real wages.

For the sake of simplicity, assume that there are two municipalities in the economy, municipality A and municipality B, which are initially in their competitive autarchy equilibria. In a strict sense, the setting describes a two-municipality economy, but the setting may also be interpreted so that municipality A is connected to the rest of the economy, and that municipality B presents the average of all the other municipalities. In any case, free mobility connects the local labour markets with each other into an inter-municipality labour market. In these circumstances both inter- and intra-municipal market conditions are determined by the market mechanism of the competitive inter-municipality labour market.

Figure 2 below shows a comparative-static presentation of the effects of migration. To add analytical power and presentational clarity, the conventional migration model of the literature of regional and urban economics (see e.g. Armstrong & Taylor, 2002, pp. 141-143; McCann, 2001, pp. 192-196) is merged with a model more familiar from the literature of international trade policy (Krugman & Obstfeld, 2000, pp. 187-253).

(Figure 2 here)

The local labour market in municipality A is presented in the left-hand-side panel and that of municipality B in the right-hand-side panel of Figure 2. The inter-municipal labour market is presented in the middle panel. The analysis starts from the initial autarchy setting and proceeds to illustrate the effects of free migration.

Given that the capital stock is constant and immobile in both municipalities, the labour demand curve is $D^A$ in municipality A and $D^B$ in municipality B. The labour demand curves are not identical, which may be due to differences in capital stocks, industrial structures, trade patterns etc. The respective labour supply curves are $S^A$ and $S^B$, and they also differ because of demographic, educational and other such differences.

The initial labour market equilibrium in municipality A is at point $e^A$ for the real wage $\omega^A$, and that in municipality B is at point $e^B$ for the real wage $\omega^B$ in Figure 2. The real wage is higher in municipality B, which brings two kinds of forces into play: on the one hand, people in municipality A are tempted to move to municipality B by the higher attainable wage $\omega^B$, and on the other hand, firms in B would like to hire labour from A for the lower wage $\omega^A$. The inter-municipal labour market in the middle panel presents these mutual interests: for real wages higher than $\omega^A$, the supply

of labour to the inter-municipal labour market from municipality A is described by schedule $S^a$, and for real wages lower than $\omega^B$, the demand for labour from the inter-municipal labour market to municipality B is described by schedule $D^b$. Technically speaking, the market supply curve $S^a$ is the horizontal difference between schedules $S^A$ and $D^A$, measuring the oversupply of labour in municipality A at wages higher than $\omega^A$. The market demand curve $D^b$ is the horizontal difference between schedules $D^B$ ja $S^B$, measuring the overdemand for labour in municipality B at wages lower than $\omega^B$.

Migration from municipality A to municipality B is guided by the inter-municipal labour market. The real wages adjust through changes in nominal wages and prices in both municipalities. Emigration from municipality A makes nominal wages in A rise because of diminishing labour supply. This is because, under constant capital stock, the productivity of labour and therefore the ability pay nominal wages rises. At the same time, since the demand for local goods and services also falls, the local price level falls. The reverse is true in municipality B: the increasing supply of labour makes the nominal wages fall and the increased demand for local goods and services makes the price level rise. The real wage rises in the out-migration municipality and falls in the in-migration municipality. The market equilibrium is found at point $\varepsilon$ in the middle panel of Figure 2, where the market demand and market supply curves intersect at the market real wage $\omega^*$. Migration ends in a stable equilibrium, where the real wages are equalised everywhere and nobody can benefit from moving to another municipality.

In the market equilibrium in Figure 2 employment in municipality A is $L^a$, and that in municipality B is $L^b$. The amount of emigration from municipality A is $L^{a1}$- $L^a$, and the corresponding immigration to municipality B is $L^b$ - $L^{b1}$. In the inter-municipal equilibrium, the total migration flow thus amounts to $L^{a1}$- $L^a$ = $L^b$ - $L^{b1}$ = $L^\varepsilon$. Because of the induced fall in real wages, the immigrants displace $L^B$ - $L^{b1}$ time units of the original working labour force in municipality B. It must be noted that

migration is measured indirectly in terms of labour time units and not directly in terms of population. The conversion of labour time units into population numbers has its problems, especially when people's labour supply decisions are taken to be continuous. However, since the measurement problem is irrelevant as to the main conclusions of this paper, it can be ignored.

Migration has welfare effects, too. The initial autarchy welfare in municipality A is measured by the area between the labour demand and labour supply curves in the left-hand panel of Figure 2. Migration makes the real wage rise to $\omega^*$ and employment diminish to $L^a$. Firms' surplus decreases by the area $\omega^A\omega^*ae^A$, of which the part $\omega^A\omega^*aa^4$ is transferred to the workers, whose total surplus is now the area $a^3\omega^*aa^2$. Migration and the induced rise in the real wage thus causes redistribution of income from the firms to those workers, who stay put in municipality A. Nevertheless, there emerges a loss of welfare experienced in municipality A that amounts to the area $a^2ae^A$.

In municipality B welfare is increased compared to the initial area between the labour demand and supply curves in the right-hand-side panel of Figure 2. This is due to two reasons. First, the increase of employment to $L^b$ fosters the firms' surplus by the area $\omega^*\omega^Be^Bb$. At $\omega^*$, the firms pay their workers less than their physical marginal product up to point $b$. Of this increase the area $\omega^*\omega^Be^Bb^1$ is taken from the original residents of municipality B, whose welfare is reduced by the fall in the real wage. Thus, taking into account the effects on the original firms and residents of B, the net increase in welfare amounts to the area $b^1e^Bb$. Second, there is the welfare gain for the immigrants from municipality A. This welfare gain can be measured by the area $a^2aa^1$ in the left-hand panel of Figure 2. After shifting the respective section of the labour supply curve with the emigrants from municipality A to municipality B, the equal welfare area $b^2b^1b$ can be found in the right-hand panel of the figure. The welfare gain emerges because the migrants would accept work for wages given by the section $a^2a^1$ (or section $b^2b$ on the dashed supply curve in the right-hand panel), but they are still

paid the market wage $\omega^*$. The total increase in welfare in municipality B is measured by the area $b^2b^1e^Bb$ in the right-hand panel.

The welfare effect of migration in the whole economy is measured by the difference of the welfare gain in municipality B, area $b^2b^1e^Bb$, and the welfare loss in municipality A, area $a^2ae^A$. Because the migrants' part of the welfare gain calculated in municipality B (area $b^2b^1b = a^2aa^1$) more than covers the welfare loss in municipality A (area $a^2ae^A$) it can be concluded that the remainder $aa^1e^A$ measures the net welfare effect for the migrants. Summing this effect with the effect on the firms in municipality B, measured by the area $b^1e^Bb$, the total welfare gain of migration amounts to the area $\omega^A\omega^B\varepsilon$ presented in the middle panel of Figure 2. The part $\omega^A\omega^*\varepsilon$ of the welfare area by definition equals the area $aa^1e^A$ in the left-hand panel, and the part $\omega^*\omega^B\varepsilon$ of the welfare area equals the area $b^1e^Bb$ in the right-hand panel.

The welfare gain of migration is occasioned by the reallocation of resources in the economy. The capital stock being immobile, migration reallocates labour force from lower-productivity municipalities to higher-productivity municipalities. In the market equilibrium the allocation of resources is efficient. This property would also persist if the labour input were immobile and the capital stock mobile, or, more generally, if both factors were mobile at the same time (Krugman & Obstfeld, 2000, pp. 70-74). Thus, migration is not a zero-sum game.

It is quite obvious from Figure 2 that migration flows and their effects depend on local labour market circumstances. At least from the viewpoint of the Finnish economy it is worthwhile to make two comments concerning the demand side of the labour market. The first comment concerns the slope of the labour demand curve in the out-migration municipality A. The steeper the labour demand curve $D^A$ compared to the rest of the economy (the more capital intensive A is in its industrial structure) the steeper is the oversupply curve $S^a$ from A to the inter-municipal labour market, and the

less responsive is migration to the real wage differentials. The market real wage also adjusts to a level higher than $\omega^*$, and the welfare effects remain smaller than those in Figure 2. Labour demand curves are steep, for example, in rural municipalities, where agriculture and/or forestry are the main branches of industry, and which export a major part of their production. In Finland, there are numerous examples of such cases.

The second comment concerns the size of out-migration municipality A. The smaller the labour force in the municipality relative to the size of the economy-wide labour market the flatter is the labour demand curve $D^b$ faced on the inter-municipal labour market. In the extreme case of an atomistic out-migration municipality the labour demand curve is horizontal at $\omega^B$. In this case migration does not affect market wages at all, and all emigrants can get a job at the constant wage rate. Wages adjust only in the out-migration municipality A, from $\omega^A$ to $\omega^B$. In general, migration is more extensive than in Figure 2, and the welfare effect, which goes totally to the emigrants, is also greater. The Finnish municipalities are typically small measured in terms of labour force.

# 3 The effects of taxation on migration and welfare

## 3.1 The initial situation

In the model, migration is driven by wage differences between municipalities. Therefore, it is reasonable to expect that labour income taxation should have effects on migration. Assuming that the municipalities are heterogenous and fiscally autonomous, differences in local tax rates are also likely to exist. It must be recalled, though, that since the model is a real wage model, differences in nominal wages net of taxes do not necessarily trigger migration. The local price level of private goods and services must also be considered, likewise the amount and quality of local public services and benefits that people receive against their tax payments. If the local public sector is efficient in the provison of services and the tax revenue is returned in full to the local taxpayers, taxation does not necessarily change the local real wage, at least not on the aggregate level.

The above model serves as a basis for the analysis of the short-term effects of taxation on migration and welfare. The starting point of the analysis is the final stage of Figure 2, where migration has stopped and the inter-municipality labour market, as well as the labour markets in municipalities A and B have settled to their short run equilibria. Municipality A is the out-migration municipality and municipality B is the in-migration municipality. The real wage is equalised in both municipalities to $\omega^*$. Figure 3 illustrates how migration has reorganised the labour market conditions in municipalities A and B and in the inter-municipal labour market.

(Figure 3 here)

The left-hand panel of Figure 3 shows that migration has resulted in a twisted labour supply schedule $S^A$ in municipality A. The lowest segment of the initial labour supply curve – from $a^3$ to $a^2$ in Figure 2 – has remained unchanged, but the rest of the schedule has changed. Because those along the segment from $a^2$ to $a^1$ have emigrated, that part of the supply curve has become vertical. Therefore the highest segment – above from $a^1$ in Figure 2 - has also shifted inwards. The height of the vertical segment of the labour supply curve depends on the magnitude of emigration, which in turn depends on the initial circumstances in municipality A relative to the rest of the economy. The steeper the initial labour demand curve the more real wages adjust in municipality A, the less emigration there is, and the shorter becomes the vertical segment. The labour demand curve $D^A$ has remained unaltered by the assumption that migration does not change the productivity of labour in the short term. The twisted labour supply curve intersects with the labour demand curve at point *a* in the left-hand panel of Figure 3, which equals point *a* in the respective panel of Figure 2.

In the right-hand panel of Figure 3 the labour supply curve in municipality B has also become twisted. This is because immigration from municipality A has shifted the curve outwards - the

segment $b^2b$ presented by the dashed line in Figure 2 has been summed horizontally to the initial labour supply curve. Below the equilibrium wage $\omega^*$ the curve has thus become flatter, and above $\omega^*$ the supply curve has shifted out by the number of emigrants. As a result, the twisted labour supply curve $S^B$ intersects with the original labour demand curve $D^B$ at point $b$ in the right-hand panel of Figure 3, equalling point $b$ in the corresponding panel in Figure 2.

In the inter-municipal labour market in the middle panel of Figure 3 all possible overdemand and oversupply curves start from the same point on the vertical axis. The demand and supply curves $D^a$ and $S^a$ are derived from the overdemand and oversupply in municipality A, and the curves $D^b$ and $S^b$ are derived respectively from municipality B. For the equilibrium real wage $\omega^*$ there is no oversupply or overdemand, and thus there is no other possible market wage such that would balance overdemand and oversupply. Therefore, the market equilibrium is stable.

### 3.2 Taxation in an out-migration municipality

Take the setting of Figure 3 as a benchmark and examine what happens when taxes are levied on labour incomes in the out-migration municipality A. An alternative interpretation is that taxes are initially equal everywhere, and municipality A raises its taxes above the rates that exist elsewhere. For the sake of simplicity, assume that the tax is a constant unit tax on labour time units (Cullis & Jones, 1998, p. 159-189). The comparative-static analysis of the effects of taxation is presented in Figure 4 for two tax alternative rates, $t$ and $t'$.

(Figure 4 here)

Examine first the effects of a tax of rate $t$ per labour time unit. The tax wedge splits the labour demand curve in municipality A into two parts, the original curve $D^A$, which determines the gross wage that the firms are able to pay, and the curve $D^A_t$, which determines the net wage, according to

which people optimise on their time use. The tax wedge *t* occurs on the vertical segment of the labour supply curve, which means that the gross wage remains at $\omega^*$, and labour demand remains at $L^a$. The net wage is $\omega^A_t$, for which labour supply also remains at $L^a$. Because the tax *t* does not affect labour demand and labour supply in municipality A, there are no effects on overdemand and oversupply on the inter-municipal labour market. Therefore, taxation affects neither migration nor the circumstances in municipality B. Nor are there any welfare effects.

The only implication of taxation is that the tax revenue in municipality A, measured by the area $\omega^A_t\omega^*aa^3$ in the left-hand panel of Figure 4, is collected wholly from the worker-residents of municipality A. This is to say that the net wage is lower in municipality A than in municipality B. Note that this does not necessarily imply a difference in the real wages – the equality of the real wages is maintained in a fair system, in which the local government provides services and transfers against the tax payments so that the taxes instantly revert to the taxpayers. But if the system is not fair, for example if the services and transfers are provided for the benefit of the firm sector, too, then taxation redistributes income from the workers to the firms, and the real wage in municipality A remains lower than that in municipality B. The system may also become unfair because of high transaction costs, ienefficiency in the local public sector etc. But in any case, the result is that tax-induced differences in real wages do not trigger inter-municipal migration in the short term.

The result that taxation does not affect migration even if it induces differences in real wages is due to the fact that the tax wedge *t* is on the vertical segment of the labour supply curve. This is to say that all worker-residents of municipality A get higher wages than the minimum acceptable. Under a fair system, this gives them extra surplus measured by the area $\omega^A_2\omega^*aa^2$. Since the tax revenue $\omega^A_t\omega^*aa^3$ is only a fraction of the extra surplus, even an unfair tax system is accepted without voting with the feet. The tax wedge can thus be as wide as the vertical segment without affecting migration and welfare.

The vertical segment of the labour supply curve constitutes a threshold for moving, which gives some degrees of freedom not only in the design of the local tax system, but also in the private wage-setting. Because there is no threat of emigration, the firms in municipality A may eventually try to shift the workers' extra surplus to their own profits. Nevertheless, since the main results of the short-term analysis do not depend on the issue, assume for simplicity that it is the workers who benefit from higher than required wages.

The above findings suggest that taxation has effects on migration and welfare only if the tax wedge is wider than the vertical segment of the labour supply curve in municipality A. Consider a tax of rate *t'* in Figure 4. Now the tax wedge induces adjustment in the gross wage along the demand curve $D^A$ to $\omega^A$, and in the net wage to $\omega^A_{t'}$ according to the point of intersection of the net of tax demand curve $D^A_{t'}$ and the supply curve $S^A$. Labour supply in municipality A falls by $L^a$-$L^{a5}$ due to the fall in the net wage below the reservation level $\omega^A_2$. The higher wage level $w^*$ in municipality B attracts these worker-residents, and new oversupply on the inter-municipal labour market emerges above wage $\omega^A_{t'}$. This is presented in the middle panel of Figure 4 by the outward shift of the labour supply curve from $S^a$ to $S^a_{t'}$.

The market equilibrium is determined on the inter-municipal labour market. Migration from municipality A to municipality B continues until labour supply $S^a_{t'}$ and labour demand $D^b$ meet at point *e* in the middle panel of Figure 4. This induces adjustment in the market wage to $\omega^e$. Immigration increases labour supply in municipality B and causes wages to fall to $\omega^e$ in the right-hand panel of Figure 4. Due to the fall in wages some of the original worker-residents of municipality B, namely $L^b - L^{b2}$, remain out of work, but new immigrants from municipality A, namely $L^a - L^{a5}$, increase the working labour force in municipality B to $L^{b3}$. At the other end, in

municipality A, the market wage $\omega^f$ is not operational – the local market conditions determine the gross wage as $\omega^A$, which is higher than $\omega^f$, and the net wage as $\omega^A_{t'}$, which is lower than $\omega^f$.

Now the tax-induced migration also has welfare effects. In municipality A welfare decreases by the area $a^4a^5aa^2$, of which the part $a^7a^5a$ is due to declining profits in the firm sector and the part $a^4a^7aa^2$ is due to lost labour income of the worker-residents. In municipality B the increase in labour supply and the consequent decrease in wages increase the profits of the firm sector by the area $\omega^f\omega^*bb^3$, and decrease the labour incomes of the original worker-residents by the area $\omega^f\omega^*bb^2$. The net benefit from this effect is measured by the area $b^2bb^3$. In addition, the new immigrants bring with them welfare measured by the area $a^4a^6a^8a^2$. The total welfare effect in municipality B is presented in the middle panel of Figure 4 by the area $\omega^A_{t'}\omega^*ee^1$ between the labour demand and supply curves. Carried to the left-hand panel of the figure, the welfare area equals $a^4a^7a^8a^2$. It can easily be seen from the left-hand panel that the welfare gain in municipality B does not fully compensate the welfare loss in municipality A – the net welfare loss is measured by the area $a^7a^5aa^8$ in the left-hand panel of Figure 4. This is the welfare cost to the whole economy due to the market distortion and the consequent inefficiency in resource allocation caused by taxation in one municipality. To sum up, by levying heavy taxes on labour income, the out-migration municipality induces a loss in its original welfare, of which part is transferred to the benefit of the in-migration municipalies while the remainder constitutes a net welfare loss to the whole economy.

The high tax *t'* creates an inter-municipal difference in workers' real wages, which is not equalised even if the local public policy is essentially fair. This can easily be seen in Figure 4. In the left-hand panel of the figure, the tax income in municipality A is measured by the area $\omega^A_{t'}\omega^Aa^5a^4$, of which the part $\omega^*\omega^Aa^5a^7$ goes to the firm sector and the part $\omega^A_{t'}\omega^*a^7a^4$ goes to the worker-residents. Refunding the tax income according to this incidence makes $\omega^*$ the real wage of the worker-

residents, which is higher than that of the worker-residents in municipality B. Equalisation of the workers' real wages necessitates active redistribution from the worker-residents to the firm sector in municipality A, and vice versa in municipality B. Such a policy is highly unlikely to emerge because, under the stable inter-municipal market conditions, there is neither market threat nor any efficiency reasons to cause it.

Figures 2-4 also facilitate a rough sensitivity analysis of the effects of taxation. First, the fact that taxation has effects only if the tax wedge is wider than the vertical segment of the labour supply curve in A turns attention to the determinants of the width of the segment. As argued above, the vertical segment remains short in municipalities in which the labour demand curve is comparatively steep, that is in municipalities with a relatively small and capital intensive industrial structure. In other words, small and capital intensive municipalities have a narrower margin to set their rax rates in order to be safe from the disadvantages of taxation than large municipalities with more diverse production structures. This is quite intuitive. On the other hand, however, Figure 4 shows that when the tax wedge is wider than the vertical segment of the labour supply curve, the implications are reversed: the steeper the labour demand curve in the out-migration municipality compared to the rest of the economy the smaller the effects of taxation. This is to say that, in this respect, small capital intensive municipalities are less vulnerable to taxation than large diverse ones.

The latter result is more counter-intuitive than the former, but it can be easily verified by elaborating Figure 4. First, assuming that municipality A is of atomistic size in the economy is to say that the market demand curve $D$ in the inter-municipal labour market is vertical at $\omega^*$. Then the welfare of the emigrants does not permeate that of the firm sector in municipality B, and the welfare loss of the economy remains smaller by the area $a^7a^5a$ in the left-hand panel of Figure 4. Second, assuming that municipality A is capital intensive in its production is to say that its labour demand curve is steeper than $D^A$ in the left-hand panel of Figure 4. Levying tax $t'$ in these circumstances would cause the

internal gross wage to adjust more and the internal net wage to adjust less than in Figure 4. The local labour market equilibrium would be reached between points $a^4$ and $a^2$ along the labour supply curve. As a result, emigration and the welfare effects would be smaller than those presented in Figure 4. The result is reasoned by the fact that the steepening of the labour demand curve changes the incidence of taxation. A capital intensive production sector is able to carry a heavy burden of taxation without noticeable effects on employment, and the consequently small reduction in net wages induces only modest migration. An important conclusion from the Finnish perspective is that taxation in a declining small, capital-intensive and unilateral municipality, such as the Finnish agricultural/forestal municipalities commonly are, has only minor, if any, effect on migration and welfare.

### 3.3 The case of an in-migration municipality

The effects of taxation must also be examined from the point of view of in-migration municipalities, because it is possible that these municipalities also experience short-term pressures for tax increases due to the need to construct infrastructure and services for new residents. Recent empirical findings suggest that, due to various real-world features, migration may induce fiscal problems both in out- and in-migration municipalities (Kallio et al., 2001). Figure 5 presents the analysis of the effects of labour income taxation imposed in an in-migration municipality.

(Figure 5 here)

In the right-hand panel of Figure 5, tax $t$ levied in municipality B creates a tax wedge between gross and net real wages, and splits the labour demand curve into $D^B$ and $D^B_t$. Because the labour supply curve, in spite of its twisted shape, moves everywhere upwards, even a small tax wedge shifts the labour supply decison backwards to $b^4$ along the flat segment of the labour supply curve, and labour demand decision backwards to $b^5$ along the original labour demand curve. Of the worker-residents of

municipality B, $L^{b5}$ continue to work in B, and the segment between points $b^4$ and $b$ of the labour supply curve is oversupply that moves to the inter-municipal labour market. The oversupply together with the original supply curve $S^b$ sum to a new labour supply curve $S^b{}_t$ in the middle panel.

The inter-municipal labour market equilibrium is reached at the intersection of $D^a$ and $S^b{}_t$ at the market wage $\omega^e$ in the middle panel of Figure 5. Migration from municipality B to municipality A is $L^{b8}$-$L^{b5}$, and the real wage in municipality A falls due to increasing labour supply to $\omega^e$ in the left-hand panel. This is to say that labour income taxation in municipality B turns the migration flow towards municipality A – the original in-migration municipality becomes an out-migration municipality. Consequently, a vertical segment in the labour supply curve now appears.

Taxation also has immediate welfare effects. The welfare loss of the out-migration municipality B amounts to the area $b^4b^5b$ in the right-hand panel of Figure 5. In in-migration municipality A welfare increases, because the immigrants carry in a welfare gain measured by the area $b^4b^6b^8$, and the firm sector in A benefits from the fall in the real wages by the area $a^{10}aa^9$ in the left-hand panel. The sum of these areas, which amounts to $\omega^B{}_t\omega^*e$ in the middle panel and to $b^4b^7b^8$ in the right-hand panel is the amount of social welfare that municipality B transfers to municipality A by its own tax policy. Calculated in the right-hand panel of the figure, the net welfare loss caused to the whole economy is measured by the area $b^7b^5bb^8$.

The tax income in municipalty B is given by the area $\omega^B{}_t\omega^Bb^5b^4$ in the right-hand panel of Figure 5. The tax incidence for the firm sector is $\omega^*\omega^Bb^5b^7$, and that to the worker-residents is $\omega^B{}_t\omega^*b^7b^4$. Just as with high taxes of rate *t'* in the above case of taxation in municipality A, the real wages are not equal between municipalities in the market equilibrium, and equalisation cannot be achieved even by fair public provision that accords with the tax incidence. Redistribution from the firm sector to the

worker-residents in municipality B would be needed to make real wages equal. Again, this would be a highly improbable and also unnecessary policy.

For a sensitivity analysis of the effects of taxation, note that the upwards-sloping labour supply curve states that even a small change in taxation causes immediate effects on migration and welfare. The first conclusion is that the original in-migration municipalities are more vulnerable to tax effects than the original out-migration municipalities. This is because the in-migration municipalities lack that kind of a migration threshold (the vertical segment in the labour supply curve) that exists in out-migration municipalities. Second, the effects depend again on the relative size and relative capital intensitivity of the municipality. As depicted in Figure 5, the reduction of labour supply is large ($L^{b} - L^{b5}$), but migration ($L^{b8} - L^{b5}$) is still modest. This is because of extensive wage adjustment on the inter-municipal labour market, where the labour demand curve is quite steep while the labour supply curve is rather flat. The slopes depend on the relative size of B (which affects the slope of the labour demand curve faced by B on the inter-municipal labour market) and B's relative capital-intensity (which affects the slope of the labour oversupply curve from B to the inter-municipal labour market).

As to the role of relative size, it is obvious in Figure 5 that the migration effect is the greater the flatter is the labour demand curve on the inter-municipal labour market, that is, the smaller municipality B is relative to the rest of the economy. If municipality B were of atomistic size, that is if the market demand curve was horizontal, emigration from municipality B would be $L^{b}$-$L^{b5}$, the whole welfare gain of migration $b^{4}b^{7}b$ would attach to the emigrants, and the welfare loss of the whole economy would shrink to the area $b^{7}b^{5}b$. The conclusion is that the effects of taxation on migration are the greater the smaller the original in-migration municipality is. The effect on welfare in B does not depend on its size, but as to the effects of taxation on social welfare, the conclusion about the effect of size is reversed – the smaller the municipality the smaller the effects on social welfare.

The role of relative capital intensity can be easily analysed by making the labour demand curve in municipality B steeper in the right-hand panel of Figure 5. The experiment results in the notion that, for a steeper set of $D^B$:s and $D^B_t$:s and, consequently, $S^b_t$, the tax wedge $t$ causes smaller effects on migration, market wage adjustment and welfare in the three panels of the figure. The conclusion is that the more capital intensive the industrial structure in the original in-migration municipality is, the less significant are the economic effects of taxation.

A general conclusion is that if an original in-migration municipality imposes a (higher-than-elsewhere) tax on the labour incomes of its residents, the immigration flow is reversed and the municipality becomes an out-migration municipality. Irrespective of the tax rate, taxation distorts the market solution, reinforces migration and causes welfare losses to the imposer itself and to the entire economy. However, it must be stressed that a large-scale in-migration municipality with a capital-intensive industrial structure can use taxation without notable effects on migration and welfare, which is an important finding, at least in the context of a recently industrialised country like Finland.

## 4 Conclusions

The paper reconsiders the economic wisdom regarding the short-term effects of labour income taxation on inter-municipal migration and welfare. The starting point is the intuitively appealing conception that, under free mobility, labour income taxation induces people to avoid taxes by migration. A common first-hand interpretation of this intuition is that tax-induced differences in wages should not be sustainable, and that the net emigration municipalities, which are under the most obvious pressure to raise their taxes, are doomed to a vicious circle of tax-accelerated emigration.

A closer investigation of the theory of inter-municipal migration reveals, however, that the first-hand conclusion is not so self-evident, at least not in the short run. The framework constructed for the analysis is a synthesis of the classical labour market model of migration familiar from regional and urban economics, and a model of trade policy familiar from international economics. People's residential choices in the model are reduced to optimal time-use decisions with respect to real wages attainable on the economy-wide labour market. The model shows that migration is not a zero-sum game. The trade policy component of the model yields further intuition on the role of inter-municipal tax differences, which cause inter-municipal differences in real wages, at least in the short term.

In its simplicity, the framework appears to be a powerful tool for analysis, and it yields some interesting results, which are also quite general in nature. The model shows, for example, that there is an asymmetry in the effects of taxation depending on whether the municipality is originally an out-migration or an in-migration municipality. An out-migration municipality can in the short term levy taxes without affecting migration and welfare. The tax must be decidedly high in order to trigger such effects. On the other hand, even a small tax increase in an in-migration municipality reverses the migration flow and causes both local and economy-wide welfare losses.

Taxation in an out-migration municipality has no effects because migration incorporates a vertical segment in the local labour supply curve. The vertical segment of the supply curve constitutes a migration threshold, which is to say that the reservation wage of the remaining worker-residents is considerably lower than the current market wage. This may be so because of professional, educational and other such personal factors determining the workers' productivity in terms of nominal wages and the prices of their specific consumption bundles. The latter explanation is of particular importance in the present model, because the price system is assumed to reflect all economic and non-economic factors of migration. Therefore, the prices include all kinds of individual transaction costs and externalities caused by owner-occupancy of housing and other real

estate value, marital status, age, social networks etc. Thus, both mobility and individually perceived real wages can vary considerably even among people of the same profession. On the aggregate, the height of the migration threshold depends on the local labour market conditions in relation to the rest of the economy.

In in-migration municipalities labour supply curves do not bend vertically and migration thresholds thus do not exist. In fact, migration makes the supply curve even flatter to include the labour supplies of both the original worker-residents and the more recent immigrants. Migration is therefore sensitive to taxation, and levying a labour income tax immediately transforms the municipality from an in-migration municipality to an out-migration municipality. That migration is sensitive to taxation in in-migration municipalities is quite reasonable, because taxation exerts a first-time incentive for emigration over the original residents, and a second-time incentive over the recent immigrants. Emigration consists of the most mobile of the original residents, and of those recent immigrants, who have not rooted very deeply in their new location of residence.

A general result of the analysis is that the short-run effects of labour income taxation on migration and welfare are less significant than what is commonly thought. Small tax increases in out-migration municipalities have no effects at all. Even higher taxes have only minor effects, when imposed by small and capital-intensive municipalities, to which category most of the declining rural municipalities belong, at least in Finland. On the other end, taxes imposed by big capital intensive in-migration municipalities have no significant effects on migration either. There may thus be ample room for inter-municipal tax differences in the Finnish economy without any threat that taxation would accelerate migration and concentration of the population.

Some critical points must be noted, however. First, the above results concern only the short-term comparative static effects of taxation. In the longer term migration concerns economic growth, too.

Allowing for variability and mobility of capital makes the effects much more complicated. The well-known corvergence hypothesis postulates that the local differences should be equalised in the long term by the dynamics of growth. Investigation of the long-term effects of taxation on factor mobility and welfare requires a macroeconomic general equilibrium framework.

Second, the migration threshold states that the emigrants have a high reservation wage compared to those who stay put in their original locations. In the present model based on the classical labour market theory this can be to some extent explained by differences in personal prices, but it is undeniable that some emigrants may be also highly skilled professionals with very high nominal wages. If so, emigration also reflects outflow of skilled labour, which is bound to have implications on the long-term prospects of the out-migration municipality. More sophisticated approaches in this respect are for example human capital theory, which includes life-cycle considerations on the costs and benefits of migration (Armstrong & Taylor, 2001, pp. 153-165) and segmented labour market theory, which allows for separate treatment of workers of different ages, education, profession, marital status etc. (Taubman & Wachter, 1986).

Third, as the present model is based on the competitive paradigm, it is unable to explain the role of agglomeration economies and diseconomies in inter-municipal evolution. Externalities, economies of scale and imperfect competition should be modelled in order to understand the endogenous formation of agglomerations. This aspect seems to be especially important in the present new-economy surroundings. (Fujita & Thisse, 2002, pp. 12-20).

Fourth, the treatment of taxation in the model is very simplistic. The unit tax system is unrealistic, and redistribution between taxpayers is not considered. The tax competition hypothesis states that free migration, where the net payers migrate to municipalities with lower tax rates and net benefitters migrate to municipalities with higher transfers, makes policy differences between municipalities

unsustainable in the long term. The analysis of this paper does yield a proper test of the tax competition hypothesis.

And fifth, in the classical labour market model, welfare differentials are introduced through the exogenous market parameters of the budget constraint assuming that all relevant factors of migration are treated by the market mechanism and reflected by market prices. In reality, however, non-economic factors also matter and notable externalities exist. The market prices cannot fully reflect the variety and quality of goods and services, the quality of residential environment, the inconveniences caused by pollution and crime etc. For a more comprehensive picture of migration, inter-municipal differences in the endogenous choice variables should also be accounted for. The theory of clubs provides a suitable framework in this respect (Cullis & Jones, 1998, pp. 294-300).

***References:***


Armstrong, H. & J. Taylor (2002), *Regional Economics and Policy*, 3rd ed., (Oxford: Blackwell Publishers).

Burda, M. & C. Wyplosz, (1997), *Macroeconomics, a European text*, 2nd ed., (New York: Oxford University Press).

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

Fisher, R. & R. Wassmer (2000), *Economic Influences on the Structure of Local Government in U.S. Metropolitan Areas,* in: Wassmer, R. (ed.), Readings in Urban Economics, issues and Public Policy, (Malden: Blackwell Publishers).

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

Kallio, O., P. Meklin, P. Siitonen & J-P. Martikainen (2001), *Muuttoliike ja kuntatalous,* (Helsinki: Suomen Kuntaliitto).

Kari, S. & H. Lönnqvist (2001), *Poimintoja taloustieteellisestä verokilpailututkimuksesta,* in: R. Hjerppe, S. Kari & H. Lönnqvist (toim.), Verokilpailu ja verotuksen kehittämistarpeet, VATT-Keskustelualoitteita, 261, (Helsinki: Valtion taloudellinen tutkimuskeskus).

Krugman, P. & M. Obstfeld (2000), *International Economics, Theory and Policy*, 5th ed., (Reading: Addison-Wesley Publishing Company).

Laurila, H. (2000), Urban Governance – kaupunkitalouden tehokas hallinta, *Finnish Local government Studies, 1/2000.*

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

Moisio, A. (2001), *On local government spending and taxation behaviour – effect of population size and economic condition,* VATT-Keskustelualoitteita, 257, (Helsinki: Valtion taloudellinen tutkimuskeskus).

Pekkala, S. & A. Kangasharju (2002), *Adjustment to regional labour market shocks*, VATT-Keskustelualoitteita, 274, (Helsinki: Valtion taloudellinen tutkimuskeskus).

Taubman, P. & M. Wachter (1986), Segmented Labor Markets, in: O. Ashenfelter & R. Lyard (eds.), *Handbook of Labor Economics*. (Amsterdam: Elsevier Science Publishers).

Figure 1: Equilibrium on the labour market

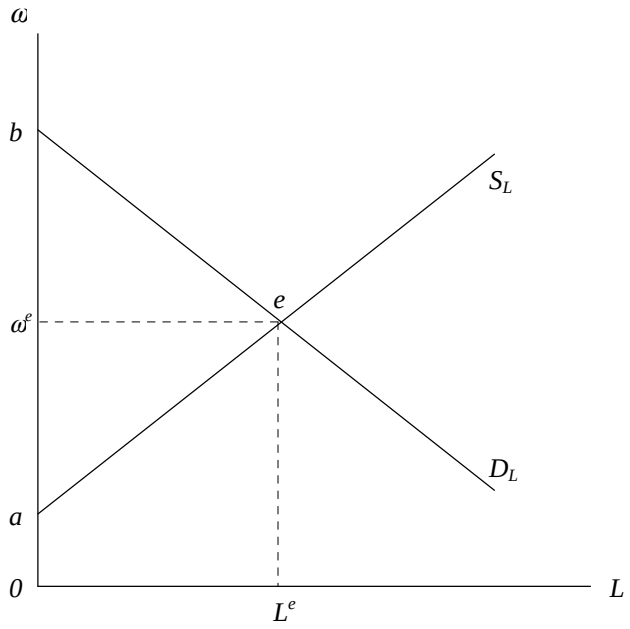


Figure 2: Migration on the labour market

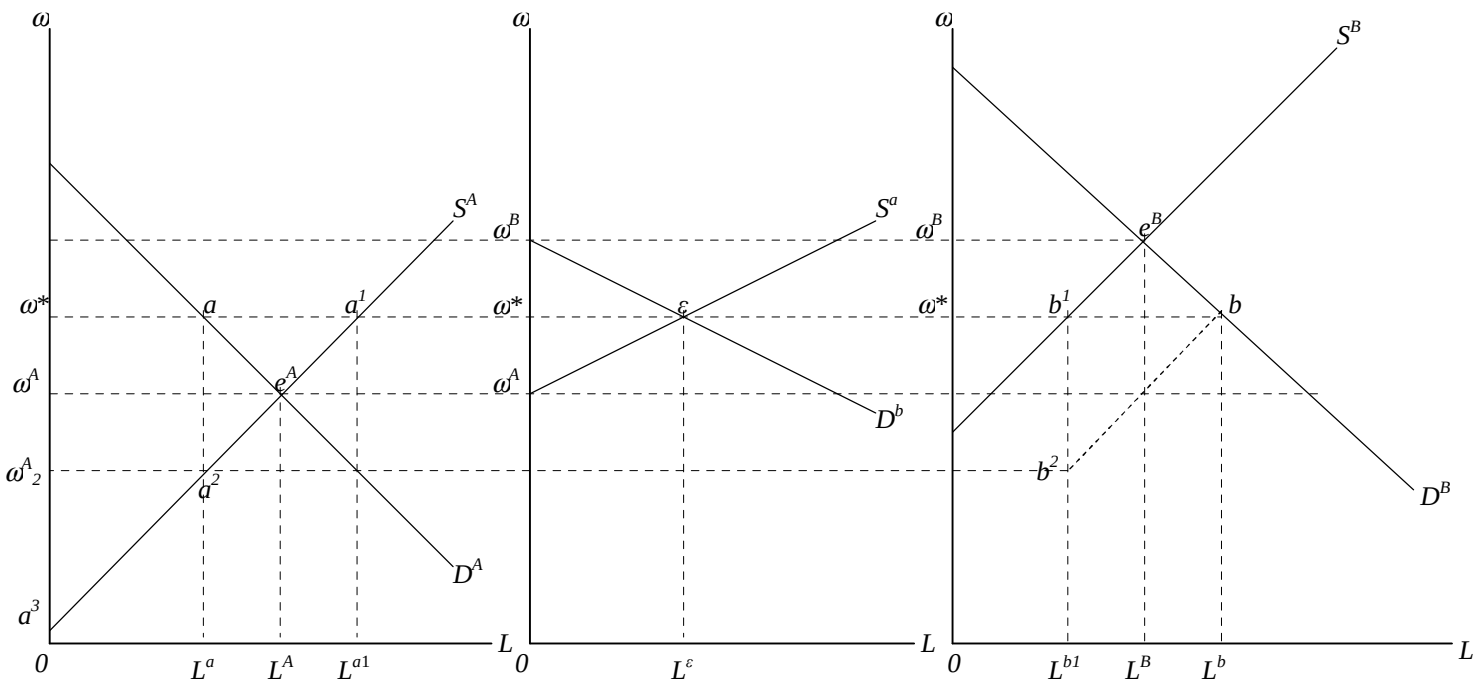


Figure 3: Post-migration equilibrium on the labour markets

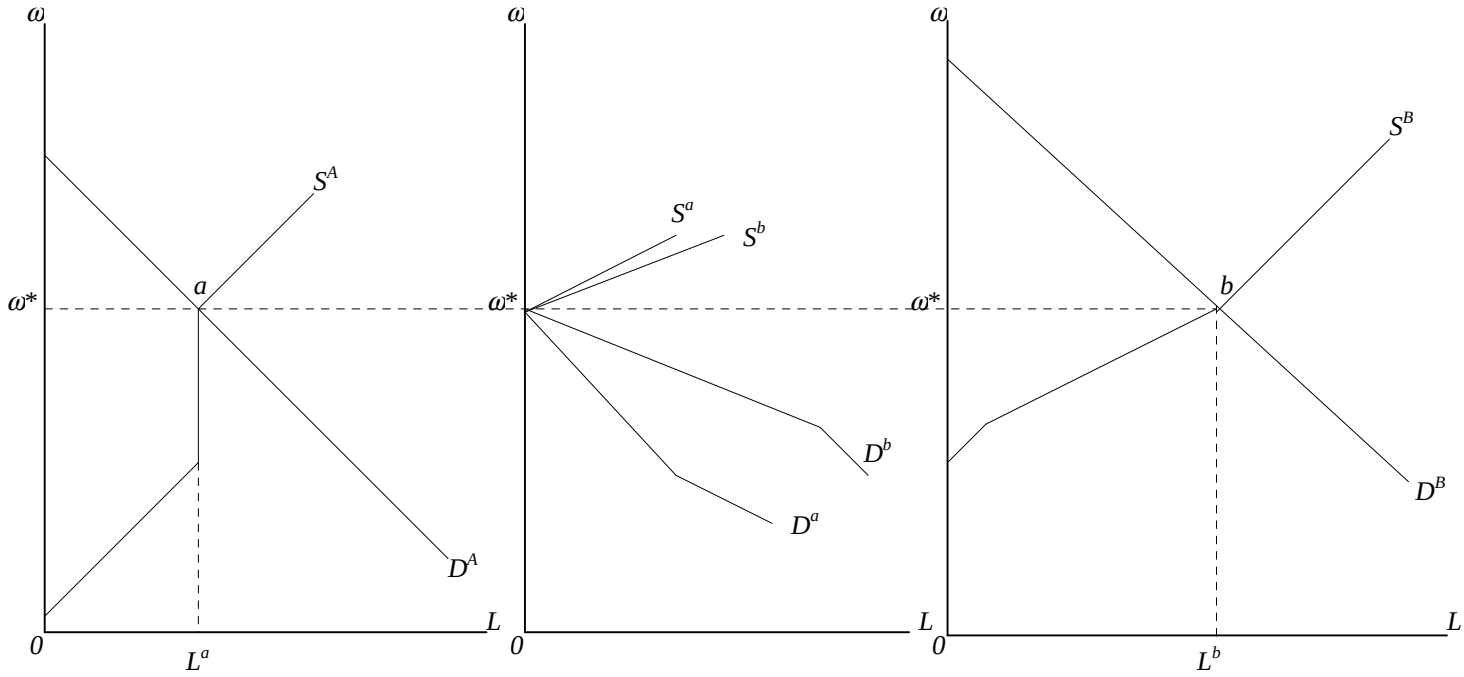


Figure 4: Effects of taxation levied in a net emigration municipality

Municipality A: Inter-municipal labour market: Municipality B:

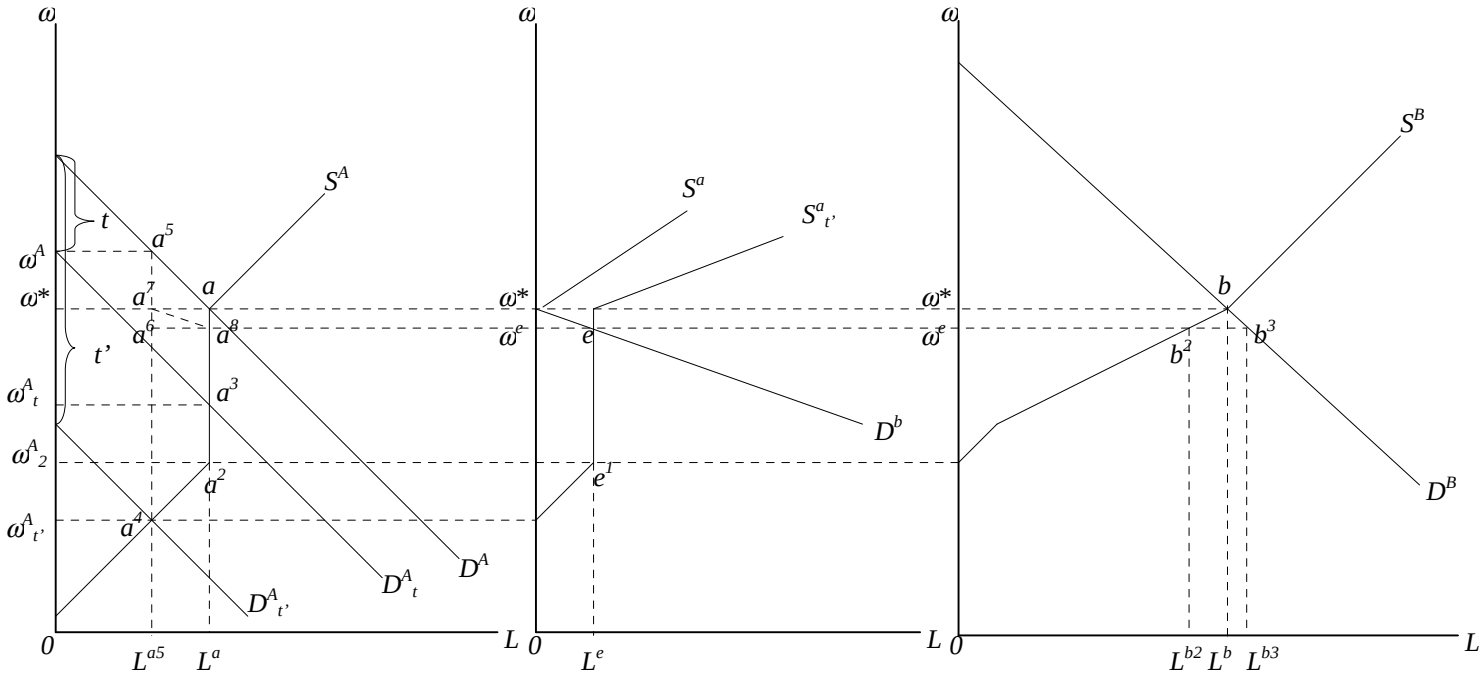
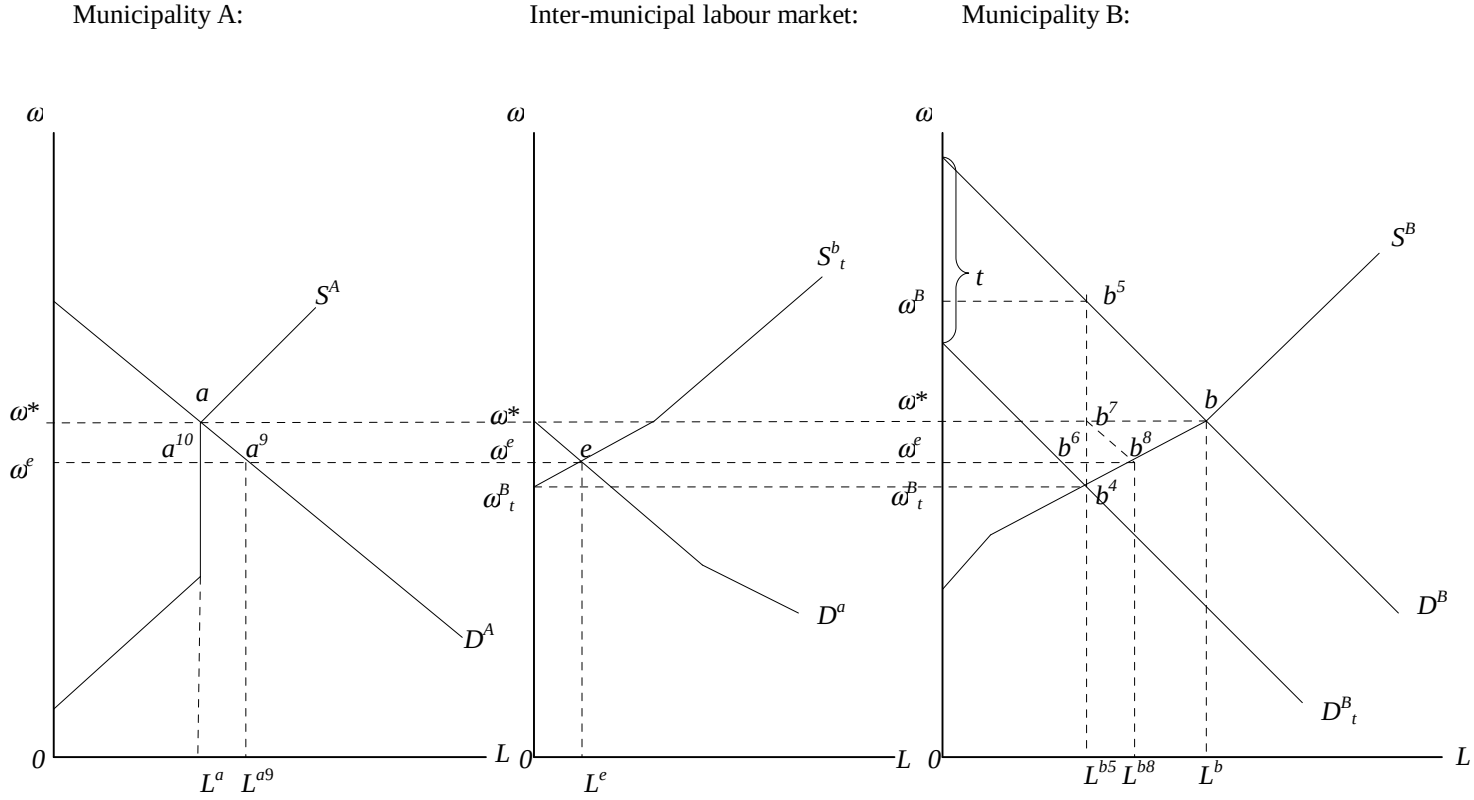


Figure 5: Effects of taxation levied in a net immigration municipality



# 3. APPROACH II: THE GENERAL EQUILIBRIUM APPROACH

## Essay #2: The Macroeconomic Model of Migration

***Abstract:*** *A neo-Keynesian macroeconomic model is constructed for the analysis of inter-regional factor migration. In the framework, mobility of both labour and capital and the consequent adjustments to the economy-wide market conditions can be treated very explicitly from the macroeconomic point of view. Four types of market adjustment are specified, namely price adjustment, nominal wage adjustment, opposite adjustment and parallel adjustment. An important finding is that the effects of labour migration on local interest rates depend on the particular type of the adjustment path. The substitution between capital movements and labour migration is studied, and the feasibility of an efficient market solution of inter-locality factor allocation is examined. The conclusion is that, from a theoretical point of view, an interior solution is not in general granted.*

*Key words: factor movements, parallel adjustment, spatial resource allocation*
*JEL classification: 931*

### 1 Introduction

The classic theory of inter-regional migration is based on a labour market application of the traditional trade doctrine. In the classic model, migration is motivated by inter-regional differences in real wages (Isard, 1975, p. 172-175; McCann, 2001, p. 191-197). Under the assumption of perfect wage flexibility, real wages adjust so that the inter-regional labour market clears. The classic labour market model is a simple but powerful device for comparative static analysis of the complex set of factors and effects of migration.

The labour market model is a reasonable framework for migration analysis for several reasons. First, work related issues constitute a major part of the determinants of average people's location decisions in practice. Second, the connection between work and residence is beyond question and the concepts of labour market area and residential area match accurately enough. To generalise even further, the concept of a local market area can be appropriately applied. And third, the setting is convenient because the real wage is a common approximation of individual welfare. This is to say

that the migrants' utility functions can be omitted and only the budget constraints that determine consumption utility can be considered. The model not only facilitates a compact treatment of a complicated issue, but also utilises concepts that can be empirically observed and measured.

The message of the classic theory of labour migration is intuitive: existing real wage differences motor up systematic migration, which eventually equalises the differences between localities. The adjustment ends up with a stable and efficient market equilibrium, where nobody can benefit from relocation and social welfare is maximised. Migration thus acts as a real world reflection of the abstract idea of factor reallocation. The simplest versions of the model ignore any closer description of local evolution of nominal wages and prices. Nevertheless, the model suggests that, in the free migration equilibrium, nominal wages and prices may well differ between localities although the real wage is equalised. This is reasonable since some important prices, like the prices of housing, are very local in nature.

This paper constructs a neo-Keynesian macroeconomic elaboration of the classic model of migration, thus providing a general equilibrium perspective on the issue. The model facilitates closer scrutiny of the adjustment of real wages between and within local market areas, which are henceforth called localities. The locality can be treated as a country, a region, a city or a municipality although the interpretations of the results may be of varying relevance depending on the treatment. Moreover, the model is appropriate for the analysis of the effects of capital movements. This is to say that the interplay between labour and capital migration can also be studied, and the feasibility of a general equilibrium based on simultaneous movements of labour, capital and even tradable goods can be examined. Application of the neo-Keynesian macroeconomic model should thus be an important theoretical clarification in the field of economics of migration.

The paper proceeds as follows. Section 2 constructs the basic neo-Keynesian macroeconomic model. The supply side of the model is composed of standard neoclassical components, treated in the classical regime. The demand side is determined by a Keynesian IS-LM setting with emphasis on local determinants of consumption demand. Section 3 provides a graphical illustration of the equilibrium properties of the model. In Sections 4 and 5, inter-locality labour migration is introduced and the migration equilibrium is examined from the viewpoints of small and larger localities. Sections 6 and 7 report the implications of capital movements and examine the existence of an interior market solution of spatial factor allocation. Section 8 concludes the findings.

## 2 The model

The model is derived from the basic neo-Keynesian macroeconomic model (Brown & Jackson, 1978; Burda & Wyplosz, 1997; Heijdra & van der Ploeg, 2002; Copeland, 2005). The model is operated in its classical regime. That is, the simple perfect foresight hypothesis is assumed to hold. The elementary results would be the same under the more modern rational expectations hypothesis.

Start by constructing the supply side of the model. Real production, denoted by $q$, is given by the production function

(1) $q = f(L, \underline{K})$,

where $L$ denotes labour input and $K$ denotes capital input. The local capital stock $\underline{K}$ is taken to be constant in the short term. Therefore, production depends on the amount of labour, measured in labour time units. The standard assumptions concerning the production function are made, namely constant returns to scale and $f_1, f_2 > 0$, $f_{11}, f_{22} < 0$, $f_{12} = f_{21} > 0$, where the subscripts refer to first and second partial derivatives of the function in order of appearance of its arguments.

Define the short-run profits of the local firm sector as $\pi = pq - wL$, where $p$ is the price level and $w$ is the nominal wage. Recalling function (1), competitive profit maximisation with respect to labour use yields

(2) $w = pf_1$

for the demand for labour in the locality. Function (2) is written in terms of nominal wages, and states that the local price level together with the marginal product of the labour input determine local labour demand - the demand for labour equals the market value of the marginal product of labour (Krugman & Wells, 2005, p. 284-289). Following from the assumption of diminishing marginal product of labour, the demand curve is downwards sloping in $L$-$w$ space.

Labour supply is determined by consumers' utility maximisation. Normalising the total available time of the households to unity, and assuming that labour is the only source of income, the households maximise their utility $U(q,1-L)$ subject to the budget constraint $wL = pq$. Substitute the budget constraint into the utility function and solve the condition for optimal time use. This yields $w/p = U_2/U_1$, where the right-hand side is the marginal rate of substitution between leisure and consumption. In order to state the time use decision in terms of labour time supply, denote $U_2/U_1 = g(L)$, where $g(L)$ can be interpreted as households' valuation of their time endowment taking into account the dual use of it (Heijdra & van der Ploeg, 2002, p. 5-8). Under the assumption of perfect foresight, local supply of labour then reads

(3) $w = pg(L)$.

Assuming that the substitution effects dominate the income effects in supplying labour time is to say that the marginal (opportunity) cost of labour time is positive, $g' > 0$. The labour supply function (3) is therefore upwards sloping in *L*-*w* space.

Given that the capital stock is constant, the local labour market conditions together with the production function determine the equilibrium output, or the aggregate supply, on the local goods market. Under perfect foresight, the derivation of the aggregate supply function, or the AS curve in *q*-*p* space is straightforward. Examination of the labour market using functions (2) and (3) reveals that, since the price parameters cancel out in the equilibrium, aggregate supply is invariant to the price level. This classical result is that, under perfect flexibility of prices and nominal wages, the aggregate supply (AS) curve is vertical in *q*-*p* space, and its horizontal position is determined by the production function, given the exogenous capital stock and the endogenous labour market solution.

For further use, however, it is instructive to give an analytic derivation of the AS curve. Totally differentiate (2) and (3), write in terms of relative changes, solve both for *dL/L*, and equate. The labour market equilibrium condition is

$$(4)\quad \frac{dw}{w} - \frac{dp}{p} = \frac{1}{\beta^S + \beta^D}\frac{dK}{K},$$

where $\beta^D = -f_1/Lf_{11} > 0$ and $\beta^S = g/Lg' > 0$ denote the real wage elasticities of labour demand and labour supply. Condition (4) states that exogenous accumulation of capital improves the real wage. To put it the other way round, the equation states that if the real wage should improve for exogenous reasons, then the increase in capital depicted should be needed to restore the labour market equilibrium.

Totally differentiate the labour demand equation (2), write in terms of relative changes, and use (4) to substitute for the change in the real wage. This yields

$$\frac{dL}{L} = \frac{\beta^S}{\beta^S + \beta^D}\frac{dK}{K}.$$

Substitute this into the totally differentiated production function (expressed in terms of relative changes), and obtain

$$\text{(AS)}\ \frac{dq}{q} = \frac{\beta^S + (1-\alpha)\beta^D}{\beta^S + \beta^D}\frac{dK}{K},$$

where $\alpha$ is used to denote the income share of labour, and $(1-\alpha)$ therefore denotes the income share of capital. The expression (AS) is the standard aggregate supply function in the classical regime of the neo-Keynesian macroeconomic model (Heijdra & van der Ploeg, 2002, p. 10). Alternatively, the aggregate supply function can be written, after manipulation and substitution of expression (4), as

$$\text{(AS')}\ \frac{dq}{q} = \left[\beta^S + (1-\alpha)\beta^D\right]\left(\frac{dw}{w} - \frac{dp}{p}\right).$$

Expressions (AS) and (AS') together state that as far as the real wage remains unchanged, production is invariant to the price level. After an exogenous accumulation of capital, the labour market equilibrium occurs at a higher real wage, and production is increased accordingly. The same also holds after an exogenous real wage increase, but only if capital accumulation supports the labour market equilibrium in the manner given by expression (4).

The demand side conditions of the local economy are determined by the relevant IS-LM model. The equilibrium of the real side of the economy, the IS curve, is given by

(5) $q = c(q^D) + i(r)$,

where $c$ denotes consumption determined by consumers' disposable income $q^D$ and $i$ denotes investment determined by the market interest rate $r$. By assumption, $c' > 0$ and $i' < 0$. Recalling the budget constraint, consumers' income is $q^D = \omega L$, where $\omega = w/p$ is the real wage. Totally differentiate (5), use the budget constraint, manipulate and obtain

$$(6) \quad dq = c' q^D \left( \frac{dw}{w} - \frac{dp}{p} + \frac{dL}{L} \right) + i' dr,$$

where

$$\left( \frac{dw}{w} - \frac{dp}{p} \right) + \frac{dL}{L} = \frac{dq^D}{q^D}$$

describes the generation of labour income along the upward sloping labour supply curve. For example, an increase in the real wage (the positive sign for the bracketed term on the left-hand side) induces increased work effort (positive sign for $dL/L$), which both sum up to an increase in labour income. In a labour market equilibrium $dq^D = dq$, and the expression for the IS curve reduces to

$$(6') \quad dq = \frac{i'}{1 - c'} dr.$$

Expressions (6) and (6') state that, since $\partial r/\partial q < 0$, the IS curve is declining in *q*-*r* space. Furthermore, exogenous shocks such that upset the labour market equilibrium shift the IS curve. Outside the equilibrium, an increase in real wages as well as that in employment shift the IS curve outwards, and vice versa.

The respective monetary equilibrium, the LM curve, of the local economy is given by the money market equation

(7) $m/p = l(q,r)$,

where the supply side is written in terms of real money, that is the amount of nominal money *m* deflated by the price level. The right-hand side of (7) depicts the demand for money, which is reasoned by two purposes: transaction use that depends on production/income *q*, and inter-temporal use that depends on market rate of return *r*, the opportunity cost of holding liquid money. By assumption, $l_1 > 0$ and $l_2 < 0$. Totally differentiating (7) and rearranging yields

(8) $$dr = \frac{1}{l_2}\frac{m}{p}\left(\frac{dm}{m} - \frac{dp}{p}\right) - \frac{l_1}{l_2}dq.$$

By (8), since $\partial r/\partial q > 0$, the LM curve is upwards sloping in *q*-*r* space for given real money supply. Furthermore, an increase in nominal money and/or a decrease in the price level increases real money *m/p* and vice versa. The LM curve shifts outwards as real money increases and inwards as real money decreases.

The derivation of aggregate demand from the IS-LM equilibrium can be carried out by substituting (8) in (6), and manipulating. The expression for the AD curve then reads

(AD) $$\frac{dq}{q} = \frac{1}{q}\frac{l_2}{l_2 + i'l_1}\left[c'q^D\left(\frac{dw}{w} - \frac{dp}{p} + \frac{dL}{L}\right) + \frac{i'}{l_2}\frac{m}{p}\left(\frac{dm}{m} - \frac{dp}{p}\right)\right].$$

In a labour market equilibrium, where $q = q^D$, (AD) reduces to

(AD') $$\frac{dq}{q} = \frac{1}{q}\frac{i'}{(1-c')l_2 + i'l_1}\frac{m}{p}\left(\frac{dm}{m} - \frac{dp}{p}\right)$$

By (AD) and (AD'), the aggregate demand curve is declining in *q*-*p* space. An increase in nominal money shifts the curve outwards and vice versa. The first term in the brackets of (AD) states that outside the labour market equilibrium, exogenous increases in employment and/or real wages shift the AD curve outwards and vice versa. Note that if employment and real wages should, for some reason, change in opposite directions, it is the net effect that determines the direction of the shift of the AD curve.

## 3 Graphical illustration of the macroeconomic equilibrium

Figure 1 illustrates the equilibrium properties of the present model. The basic figure consists of four quadrants, the labour market, specified in terms of nominal wages in the northeast quadrant, the production function in the southeast quadrant, the aggregate goods market in the southwest quadrant and the real wage in the northwest quadrant. A fifth quadrant is added at the far southwest end of the figure to incorporate the IS-LM setting of the economy.

(Figure 1)

In Figure 1, the initial equilibrium $e_0$ on the labour market is achieved by the adjustment of nominal market wages and prices to $w_0$ and $p_0$, which give $w_0/p_0 = \omega_0$ for the equilibrium real wage. The labour market equilibrium gives $L_0 = L(\omega_0)$ for labour use, which in turn yields production $q_0$ along the production function. The combination $p_0$, $q_0$ is the aggregate supply on the goods market, presented by point $E_0$ in the southwest quadrant of the figure. Since aggregate supply is invariant to the price level around the equilibrium point, it can be drawn in its full shape by the graph $AS_0$.

The $p_0$,$q_0$ combination is sustainable given that the goods market is in equilibrium. That is, aggregate supply, derived above from the labour market and production conditions must equal aggregate demand, which is derived from the IS-LM setting. The far southwest quadrant of Figure 1 presents the corresponding IS-LM equilibrium at the intersection point $\varepsilon_0$ of the downward sloping $IS_0$ curve and the upward sloping $LM_0$ curve, which gives $r_0$ for the equilibrium interest rate. The IS-LM equilibrium is assumed to yield aggregate demand matching point $E_0$ in the goods market, through which the aggregate demand curve strikes in a downward sloping manner. For later purposes, the aggregate demand curve is drawn in three hypothetical versions, $AD_0$, $AD_0'$ and $AD_0''$.

The general equilibrium nature of the model can be seen by investigating the adjustment path from the initial equilibrium to another one after an exogenous shock in the capital stock. Recall the two expressions for the AS curve, (AS) and (AS'), and consider an increase of the real wage from $\omega_0$ to $\omega_1$ such that is fully supported by an increase in the capital stock from $K_0$ to $K_1$. The real wage line rotates clockwise in the northwest quadrant and the production function shifts outwards in the southeast quadrant of Figure 1. As capital accumulation improves the productivity of labour, the labour demand curve shifts outwards from $D_0$ to $D_1$ in the northeast quadrant of the figure.

Figure 1 presents four analytically interesting benchmark cases of possible equilibrium adjustment paths. The first case is labelled *nominal wage adjustment*. In this case, the price level is fixed at $p_0$,

and the adjustment on the labour market is carried out by a rise in nominal wages from $w_0$ to $w_1$. The labour market equilibrium shifts to $e_1$, employment increases to $L_1$, and production increases to $q_1$. The vertical aggregate supply curve shifts outwards to $AS_1$. Pure nominal wage adjustment cannot, however, yield a new general equilibrium. This is evident from the fact that the respective $p_0$, $q_1$ combination at point $E_1$ in the goods market is not sustainable. This is because the aggregate demand curve is by (AD) unambiguously downward sloping and cannot therefore cross the $AS_1$ curve at $E_1$. Unless there are no exogenous shocks other than that in capital stock, at least some price adjustment must occur.

The second benchmark case is called *price adjustment*. This kind of special case happens with an aggregate demand curve like $AD_0$, backed up by $IS_0$ in the far left quadrant of Figure 1. Taking $AD_0$ and $AS_1$ as given, the goods market equilibrium at $E_2$ precludes that prices fall from $p_0$ to $p_1$. The fall in prices makes the labour demand curve shift inwards from $D_1$ to $D_2$ and the labour supply curve shift outwards from $S_0$ to $S_1$. The new labour market equilibrium is at $e_2$ with nominal wages remaining fixed to $w_0$. The fall in prices also causes the LM curve to shift outwards to $LM_1$ in the far left quadrant, where the new IS-LM equilibrium is at $\varepsilon_2$. The equilibrium interest rate is $r_1$.

Third, both nominal wages and prices may adjust simultaneously so that nominal wages rise and prices fall. This case is labelled *opposite adjustment*. This kind of adjustment is occasioned up by an aggregate demand curve such as $AD_0$', backed up by $IS_0$'. The relevant goods market equilibrium $E_3$ necessitates that the prices adjust to $p_2$, making labour demand and supply shift to $D_3$ and $S_2$ respectively. The labour market equilibrium thus settles to $e_3$ at the nominal wage level $w_2$. The IS-LM equilibrium is at $\varepsilon_3$ implying that the market interest rate is $r_2$.

Fourth, prices and nominal wages can also adjust simultaneously in the same direction so that prices fall below $p_1$ and nominal wages fall below $w_0$. Call this possibility the case of *parallel adjustment*.

This kind of adjustment is provoked by the steep aggregate demand curve $AD_0$", which induces prices to fall to $p_2$. The fall in prices, in turn, causes the labour demand and supply curves to intersect at $e_4$ so that the equilibrium nominal wage is $w_3$. The real wage has improved because the relative change in prices is greater than that in nominal wages. The market interest rate now falls to $r_3$, determined by the intersection of $IS_0$" and $LM_3$.

The main message of this analysis is that supply side shocks induced by capital inflows/outflows leave the demand side on the goods market unaffected in spite of the fact that the real wage has changed. Higher/lower real wages motivate increased/decreased labour supply along the aggregate labour supply curve, but aggregate demand is still determined by the initial, downwards sloping AD curve. Thus, the price level must decline/rise to produce equilibrium on the goods market. It is the slope of the AD curve that determines the relative adjustment in prices and nominal wages. Nominal wages may adjust in either direction or may not adjust at all, but it is for sure that prices and interest rates cannot remain unchanged.

## 4 Migration of labour

The standard way to analyse the factors and effects of migration is to construct a comparative-static setting, where the autarchy equilibrium of one locality is compared to that of the outer economy. The intuition is that migration, triggered by inter-locality differences in real wages, leads to the equalisation of these differences as real wages rise in the out-migration localities and fall in in-migration localities. After the real wages are equalised, systematic migration ends because nobody can benefit from relocating. The result is a stable and efficient market outcome.

Real wage adjustment may work through the evolution of either nominal wages or price levels. In practice, both usually adjust at the same time. Concentrate on migration alone and assume that capital remains fixed and immobile, and that there is no trade in goods and services between

localities. The labour input is the only mobile element in the economy, and the localities are assumed to constitute perfect labour market areas so that working in a locality means residing in it, and taking a job in another locality necessitates moving. Perfectly foresighted households are able to perfectly monitor possible welfare differentials. Assume also that migration is costfree so that there are no migration thresholds of any kind. Full employment, wage flexibility, competitiveness and perfect foresight throughout the economy are also assumed.

Immobility of capital and non-tradability of goods mean that interest rates and prices are determined purely locally. This implies that the presented IS-LM framework is relevant without further restrictions. The assumption best suits very large localities with autonomous goods and money markets and full interest rate flexibility.

Examine migration first from the point of view of a small locality. The local market equilibrium must fully adjust to the market conditions of the rest of the economy, and that migration flows in either direction do not affect those circumstances. Figure 2 illustrates the effects of shifting from autarchy to free migration.

(Figure 2)

In Figure 2 the initial autarchy equilibrium in the locality considered is given by the solution on the competitive labour market. The market prices and nominal wages have been settled so that the real wage is $\omega_0 = w_0/p_0$. The equilibrium occurs at point $e_0$ in the labour market, yielding the consequent outcomes in labour use, $L_0$ and production, $q_0$. The goods market equilibrium is at $E_0$, and the IS-LM equilibrium is at $\varepsilon_0$, the respective equilibrium interest rate being $r_0$.

Suppose that the real wage is higher outside the locality, denoted by $\omega^*$ in the northwest quadrant of Figure 2. Under conditions of perfect mobility of labour workers are employed in the economy-wide labour market and the local general equilibrium must adjust to this condition. Since the exogenous real wage shock is now assumed not to be supported by capital accumulation, and since the induced solutions on the labour market are not those of equilibrium, the macroeconomic adjustment now differs somewhat from that described above.

Start by examining the effects on the supply side. Fix the price level at $p_0$ and examine the adjustment simply in terms of nominal wages. The adjustment from $\omega_0$ to $\omega^*$ necessitates a rise in the nominal wage from $w_0$ to $w_1$. Reading horizontally at $w_1$, labour demand and labour supply are determined according to the original $D_0$ and $S_0$ schedules so that demand is $L_1$ and supply is $L_2$. The excess supply of labour – the amount of emigration that is necessary to restore domestic labour market balance at the constant price level – amounts to $L_2$ - $L_1$. Since the use of domestic labour falls from $L_0$ to $L_1$, domestic production falls from $q_0$ to $q_1$ and the aggregate supply curve shifts inwards from $AS_0$ to $AS_1$.

Because of the exogenous effect on real wages and employment caused by free migration, the aggregate demand conditions also change. The planned consumption expenditure of the worker/residents is determined according to the labour supply curve $S_0$, but since domestic employment is restricted to $L_1$ from the demand side, and the workers are paid according to the $D_0$ curve, the real wage improves for those who remain employed locally. Thus, the employment effect and the real wage effect have opposite signs in expressions (6) and (AD) in Section 2.

Since the AD curve sets the price level in the local economy, the final adjustment depends on the exact shift of the curve. For analytical ease, the above four definitions of types of adjustment are used again. In the first three the net of employment and real wage effects are assumed to be

negative so that the IS curve and consequently the aggregate demand curve shift (sufficiently) inwards. The possibility that the real wage dominates the employment effect so that the IS and AD curves shift outwards is commented as the fourth case. Note that the four cases are again purely hypothetical in Figure 2.

First, consider the case of *nominal wage adjustment*. Assume that the IS curve shifts from $IS_0$ to $IS_1$, and the AD curve shifts from $AD_0$ to $AD_1$. The IS-LM equilibrium shifts to $\varepsilon_1$ along the unchanged $LM_0$ curve, and the interest rate falls from $r_0$ to $r_1$ restoring the balance in the money market – money released from transaction demand shifts to inter-temporal purposes. The new goods market equilibrium is at point $E_1$ for the unchanged price level $p_0$. Nominal wages adjust to $w_1$, and the resulting emigration amounts to $L_2$ - $L_1$. With migration, it thus is possible that the local market adjustment happens solely in terms of nominal wages.

Second, consider *price adjustment*. Assume that the shift of the IS curve is from $IS_0$ to $IS_2$, and the consequent shift of the AD curve is from $AD_0$ to $AD_2$. Equilibrium on the goods market necessitates that the price level must fall from $p_0$ to $p_1$. Because of the fall in prices, the labour demand curve shifts from $D_0$ to $D_1$, and the labour supply curve shifts from $S_0$ to $S_1$ so that the new point of intersection is at $e_1$ vertically below point $e_0$ in the northeast quadrant of Figure 2. Reading at the unchanged nominal wage $w_0$, domestic labour demand is $L_1$ and labour supply is $L_2$. The excess supply of labour $L_2 - L_1$ is again emigration that balances the local labour market. Furthermore, the fall in prices increases the supply of real money and shifts the LM curve outwards from $LM_0$ to $LM_1$. The interest rate falls from $r_0$ to $r_2$.

Third, consider *opposite adjustment* of prices and nominal wages, and assume that the IS and AD curves shift respectively to $1S_3$ and $AD_3$. This causes adjustment in both prices and nominal wages, because at the price level $p_2$, set by point $E_2$ in the goods market, adjustment to $\omega^*$ in terms of

migration requires that the nominal wages rise to $w_2$. The price level $p_2$ results in the labour demand and labour supply schedules $D_2$ and $S_2$, and in the money market equilibrium schedule $LM_2$ in Figure 2. The equilibrium interest rate settles at $r_3$. As to migration and production, the solution is equal to the above cases of unilateral adjustment.

Fourth, since there is the possibility that the IS and AD curves may also shift outwards, the case of *parallel adjustment* of prices and nominal wages is also relevant. This possibility is shown in Figure 2 by dashed curves. Imagine that the AD curve shifts outwards from $AD_0$ to $AD_4$. The price level must then rise from $p_0$ to $p_3$, and the corresponding labour demand and supply curves $D_3$ and $S_3$ must intersect horizontally above $e_0$ at $e_3$. Moreover, nominal wages must rise relatively more than the price level to match $\omega^*$. Local employment, production and emigration are the same as above, but an important finding is that the market interest rate now rises from $r_0$ to $r_4$. This is because the IS curve has shifted outwards to $IS_4$ and the LM curve has shifted inwards to $LM_3$ due to the rise in the price level. Note that the figure shows that even if the AD curve does not shift at all, or if the shift is small enough (less than from $AD_0$ to $AD_1$), what will happen will be parallel adjustment.

The experiments above show that higher/lower real wages outside the locality considered drive emigration/immigration, and induce adjustment in local aggregate supply and aggregate demand conditions. The real effects on the supply side do not depend on the particular path of adjustment to the inter-locality real wage. The adjustment may occur through local prices, local nominal wages or both. The final local prices, nominal wages and interest rates are determined by the effects of emigration on local demand. The effect on local aggregate demand depends on the net of the real wage and employment effects, which in turn depends on the elasticity of labour supply and labour demand, and on the real wage difference to be adjusted to. It is important to note that local interest rates may fall or rise due to migration depending on the particular adjustment path: parallel

adjustment in prices and nominal wages turns the interest rate effects the other way round as compared to other adjustment paths.

Theoretically speaking, the path of *parallel adjustment* seems quite probable, because other kinds of adjustment paths require considerable changes in aggregate demand to start them up. Empirically speaking, however, it is arguable that considerable changes do occur – there is ample evidence that prices fall rather than rise in out-migration localities, and vice versa. This is reasonable, especially from the point of view of housing and real estate prices, which constitute a considerable proportion of local consumption expenditures, and which are determined by markets with an inelastic short-term supply side (Laakso & Loikkanen, 2004, pp. 267-280).

## 5 Equilibrium between two large localities

In practice, some localities are often big enough so that migration from/into that locality affects the conditions on the inter-locality labour market. If so, migration also has local macroeconomic implications at both ends of the flow. To analyse these implications, assume now that the economy consists of two symmetrically large localities, A and B. Suppose that real wages initially differ between the localities, and examine the effects of opening up free migration between these two localities.

Figure 3 below presents the initial autarchy situation in the two localities, the market mechanism of simultaneous adjustment, and the final allocation of the perfectly mobile labour resources. In the figure, panel (i) is a four-quadrant representation of the general equilibrium in locality A, panel (iii) is the corresponding presentation of locality B, and panel (ii) presents the inter-locality labour market, drawn in terms of real wages. In the figure, the IS-LM configurations are ignored, but the respective aggregate demand schedules are drawn.

(Figure 3)

The initial autarchy solution in locality A in panel (i) of Figure 3 is given by the labour market equilibrium at point *a*, corresponding to real wage $\omega^A$. The corresponding autarchy solution in locality B in panel (iii) is at *b* given by the real wage $\omega^B$. Free migration causes adjustment in both localities so as to match the equilibrium solution in the inter-locality labour market. Assume that the market adjustment process follows the path of *opposite adjustment* in both localities.

Consider first locality A. Since the real wage is lower in B, the workers in A are happy with their location because they cannot improve their welfare by moving. Thus, there is no motivation to emigrate. On the other hand, the firms in locality A would be willing to employ people for lower wages. Therefore, there exists over-demand for workers from B at lower real wages than $\omega^A$. There is a pull factor for immigration, which amounts to the horizontal difference between the domestic labour demand and supply curves below $w_0{}^A$ in panel (i). This excess demand is redrawn in panel (ii), where the downward sloping demand curve $D_L = D_L{}^A - S_L{}^A$ illustrates the labour demand in the inter-locality labour market.

In locality B the difference in real wages is anticipated from the opposite angle. The firm sector recruits all the workers it can for the real wage $\omega^B$, but there are workers, who would be keen to accept work for higher real wages. Therefore there exists a push factor for emigration in locality B: the over-supply of labour above the real wage $\omega^B$ constitutes the upward sloping labour supply curve $S_L = S_L{}^B - D_L{}^B$ in the inter-locality labour market.

At the moment when mobility of labour becomes free, labour demand and labour supply meet in the inter-locality labour market in panel (ii) of Figure 3. The real wage adjusts to $\omega^*$ generating the

inter-locality labour market equilibrium at point *e*. In the equilibrium, inter-locality migration is $L_e$. In locality A the adjustment means that the nominal wage falls from $w^A_0$ to $w^A_1$ and the price level rises from $p^A_0$ to $p^A_1$ in panel (i) of Figure 3. The rise in prices shifts the labour demand curve outwards to $D^{A'}$ and the labour supply curve inwards to $S^{A'}$. Reading horizontally at $w^A_1$ along the $S^{A'}$ and $D^{A'}$ curves, domestic labour supply falls from $L^A_0$ to $L^A_1$, but labour use rises to $L^A_2$. The discrepancy $L^A_2 - L^A_1 = L_e$ is bridged by immigration. Therefore, production increases from $q^A_0$ to $q^A_1$, the aggregate supply curve shifts from $AS^A_0$ to $AS^A_1$ and the aggregate demand curve shifts from $AD^A_0$ to $AD^A_1$.

In locality B the adjustment goes the other way round: the nominal wage rises from $w^B_0$ to $w^B_1$ and the price level falls from $p^B_0$ to $p^B_1$ in panel (iii) of Figure 3. The fall in prices causes the labour demand curve to shift inwards from $D^B$ to $D^{B'}$ and the labour supply curve to shift outwards from $S^B_0$ to $S^B_1$. Reading horizontally at $w^B_1$, domestic labour use falls from $L^B_0$ to $L^B_1$. The difference between labour supply and labour demand $L^B_2 - L^B_1 = L_e$ flows out in the form of emigration. As a result, domestic production diminishes from $q^B_0$ to $q^B_1$. The aggregate supply schedule shifts inwards from $AS^B_0$ to $AS^B_1$, and the aggregate demand schedule shifts inwards from $AD^B_0$ to $AD^B_1$.

It is plausible that, with free migration of labour, there exists an interior solution of residence (or spatial allocation of labour) in a multi-locality economy. Migration equalises the real wages in the two localities, but nominal wages and price levels are not necessarily equalised. For example, the price of housing usually differs considerably between in- and out-migration localities. Furthermore, recalling the message of Figure 2, local interest rates should adjust in opposite directions: the interest rate rises in locality A and falls in locality B. With immobile capital and no trade, big differences in local prices, nominal wages and interest rates may thus exist. Things may, however, be different in a more general situation, where other elements besides labour are also mobile.

## 6 Migration of labour and capital as substitutes

The analysis of the properties of the present model, as illustrated in Figure 1, follows the neoclassical intuition that capital accumulation increases employment and production, improves the real wage, and lowers the market interest rate. The analysis of labour migration, illustrated in Figure 2, supplements the intuition by showing that emigration is an alternative way of adjustment to those exogenous changes in the real wage which are not supported by capital accumulation. Figure 4 below combines these two elements and illustrates the equilibrium of simultaneous factor movements from the point of view of a small locality. The figure presents the analysis in the basic four-quadrant illustration omitting the IS-LM framework for the sake of simplicity.

(Figure 4)

The initial autarchy equilibrium of the labour market is again at point $e_0$ at the intersection of the $D_0$ and $S_0$ curves in Figure 4. The nominal wage is $w_0$, the price level is $p_0$, and the real wage is $\omega_0$. Employment is $L_0$, production is $q_0$, and the aggregate supply curve is $AS_0$. The aggregate demand curve is $AD_0$.

Let the constant market real wage in the outer economy be $\omega^*$. Read the migration potential without capital accumulation in the case of *nominal wage adjustment* at fixed $p_0$. The required adjustment in the nominal wage is from $w_0$ to $w_1$. Domestic employment diminishes to $L_1$, production falls to $q_1$, and the aggregate supply curve shifts to $AS_1$. Suppose that the initial price level $p_0$ is sustainable from the point of view of the demand side so that the aggregate demand curve shifts to $AD_1$, and the goods market equilibrium shifts to $E_1$. The respective migration potential is $L_2 - L_1$.

Next, assume exogenous capital accumulation. As the local capital stock increases from $K_0$ to $K_1$, the production function shifts outwards from $f(L,K_0)$ to $f(L,K_1)$ in the southeast quadrant, and the

real wage line is rotated clockwise from $\omega_0$ to $\omega_1$ in the northwest quadrant of Figure 4. The labour demand curve shifts outwards to $D_1$, and the autarchy labour market equilibrium shifts to $e_1$, yielding $L_3$ for equilibrium employment. In the autarchy goods market, however, the increased production would be consumed only if prices should fall to $p_1$ along $AD_0$. The fall in prices causes the labour demand curve to shift inwards to $D_2$ and the labour supply curve to shift outwards to $S_1$. The autarchy labour market equilibrium would be at $e_1$, and the relevant migration equilibrium can be read vertically from $\omega_1$ to $\omega^*$. To simplify the figure, aggregate demand curve is assumed to shift from $AD_0$ to $AD_2$ so that equilibrium migration can be read at $p_1$, horizontally along $w_0$. Migration amounts to $L_2 - L_4$, local production is $q_2$, and the goods market equilibrium is at $E_2$. Even if the final price level had settled differently, the migration effect would have been the same.

Comparison of the migration potentials without and with exogenous capital accumulation reveals that, since $L_4 > L_1$, then also $L_2 - L_1 > L_2 - L_4$. The result shows that the potential for emigration is inhibited by the increase in the domestic capital stock. This is intuitive because the capital inflow increases the productivity of labour in domestic use. An inflow of capital is clearly a substitute for emigration. Capital inflow can fully offset the need for emigration, if the impact of capital accumulation is strong enough to make the domestic real wage the same as the market wage in the outer economy, or $\omega_1 = \omega^*$ in terms of Figure 4. Moreover, the impact may even be high enough to make the domestic real wage line steeper than that of the outer economy, $\omega_1 > \omega^*$, thus the initial migration pattern is reversed and emigration turns into immigration.

## 7 Full factor mobility

In the contemporary real world, both labour and capital are increasingly mobile. The conventional wisdom, normally derived in a classic general equilibrium framework, is that an equilibrating mechanism also exists in this context (McCann, 2001, pp. 208-221; Krugman & Obstfeld, 2003, pp. 160-175). This is reasonable because differences in regional capital/labour ratios are reflected in the

marginal products of the factors causing labour and capital to start to migrate in opposite directions with consequent opposite effects on the factor rewards. Therefore, factor mobility results in equalisation of the marginal products of the factors and the capital/labour ratios between localities thus producing an interior solution of factor allocation.

Figure 5 below represents the above argument in the neo-Keynesian framework of this paper. It is assumed that the economy consists of two symmetrically large localities, between which labour, capital and consumption goods are all mobile. The figure consists of two panels, panel (i) for locality A and panel (ii) for locality B. The inter-locality labour market is omitted, but the IS-LM frameworks for the two localities are presented. In order to obtain any results, the focus is on the movements of capital and labour, and the tradability of goods is treated by fixing the price level to $p^*$ in both localities. This can be deemed to reflect the free trade equilibrium of consumption goods in the economy, and it implies, that only *nominal wage adjustment* is possible.

(Figure 5)

In Figure 5, the initial situation in locality A is given by the labour market equilibrium $a_0$ at the real wage $\omega^A$ in panel (i) of the figure. At capital stock $K^A_0$ and employment $L^A_0$, production is $q^A_0$. The initial equilibrium interest rate is $r^A_0$. The respective equilibrium in locality B in panel (ii) of the figure is described by point $b_0$ at real wage $\omega^B < \omega^A$ in the labour market. The capital stock $K^B_0$ and employment $L^B_0$ give $q^B_0$ for production. The interest rate is initially $r^B_0 > r^A_0$.

First, analyse the motives for migration. In the initial situation there is a difference in real wages in favour of locality A, $\omega^A_0 > \omega^B_0$. This accelerates migration from locality B to locality A. The IS and AD curves shift outwards in locality A and inwards in locality B. The migration effects are seen by monitoring the required adjustment in terms of nominal wages at the fixed price level $p^*$ in both

localities. In the (omitted) inter-locality labour market, the migration equilibrium is achieved by adjustment of the local nominal wages so that a common equilibrium real wage $\omega^*$ is found. This implies that, prices being fixed to $p^*$ in both localities, the nominal wages must be equalised, too. It is also obvious that employment, production and the interest rate are caused to rise in locality A and fall in locality B.

Second, examine the simultaneous effects of capital movements. In the initial situation in Figure 5, the real interest rate in locality B is higher than that in locality A, $r^B_0 > r^A_0$. The difference in the market interest rates attracts capital from locality A to locality B. Capital decreases from $K^A_0$ to $K^A_1$ in locality A, and increases from $K^B_0$ to $K^B_1$ in locality B. The real wage is caused to decrease in locality A and increase in locality B. In response, the labour demand curve shifts inwards in locality A from $D^A_0$ to $D^A_1$ and outwards in locality B from $D^B_0$ to $D^B_1$. Employment and production are caused to fall in locality A and rise in locality B, but the interest rate is further caused to rise in locality A and to fall in locality B.

In Figure 5, the simultaneous equilibrium on labour and capital markets is given by $p^*$, $\omega^*$ and $r^*$. Reading the result in terms of labour, the migration equilibrium is $L^A_1$-$L^A_2$ = $L^B_3$-$L^B_1$ (immigration to A equals emigration from B) at the equilibrium real wage $\omega^*$. Furthermore, $L^A_3$-$L^A_1$ measures the amount of immigration that the capital outflow offsets in locality A, and $L^B_1$-$L^B_2$ is the respective amount of emigration that capital inflow offsets in locality B. In the general equilibrium, production in locality A increases from $q^A_0$ to $q^A_1$, and that in locality B decreases from $q^B_0$ to $q^B_1$. The conclusion is that an interior market solution of spatial factor allocation between the two localities is feasible.

It is arguable, however, that the above conclusion does not hold in general in the present neo-Keynesian framework. First, the analysis of Figure 5 is based on the assumption that the factors are

initially drawn to opposite directions ($\omega^A_0 > \omega^B_0$ and $r^A_0 < r^B_0$, or $\omega^A_0 < \omega^B_0$ and $r^A_0 > r^B_0$). It is quite obvious from the figure that all other initial circumstances lead to corner solutions. Proper growth models with factor mobility usually apply special assumptions to avoid corner solutions (Andersson & Kuenne, 1986; Barro & Sala-i-Martin, 2004, p. 383-407). Second, the assumption that the price level is fixed at $p^*$ might also be questionable in terms of generality. If local prices should be allowed to adjust, too, then the effects on local interest rates would become dependent on the particular types of local adjustment mechanisms. The possibility of *parallel adjustment* suggests that the convergence of the interest rates would become theoretically unclear. The conclusion is that, in the present model without further restrictions and policy considerations, a stable and efficient interior solution of inter-locality factor allocation is not in general granted.

## 8 Conclusions

The main contribution of the paper is to apply the neo-Keynesian macroeconomic model to the analysis of factor migration. The virtue of the present model is threefold. First, the model takes a general equilibrium viewpoint on migration instead of the partial equilibrium perspective of the classic labour market model. Macroeconomic effects throughout the localities considered can thus be examined. Second, even compared to earlier general equilibrium presentations, the present model allows a more detailed analysis of local market adjustment to economy-wide market conditions. Simultaneous analysis of labour and goods markets is of particular importance. Third, the model facilitates simultaneous treatment of labour and capital migration. Being also quite manageable in terms of graphics, the model yields a major theoretical clarification concerning the effects of factor migration. As such, the model should serve as a sound basis for policy considerations and other such experiments.

The model is first studied around the general equilibrium. The analysis shows that supply side shocks caused by exogenous capital inflows (outflows) shift aggregate supply outwards (inwards),

but leave the demand side on the goods market unaffected in spite of the fact that real wages have also changed. Higher/lower real wages motivate increased/decreased labour supply along the aggregate labour supply curve thus constituting planned consumption along a downwards sloping aggregate demand curve. Thus, the price level must decline (rise) to restore equilibrium on the goods market. The slope of the aggregate demand curve determines the relative adjustment in prices and nominal wages. Four types of adjustment paths are specified, namely *price adjustment, nominal wage adjustment, opposite adjustment* and *parallel adjustment*. The key finding is that nominal wages may adjust in any direction or may even remain unchanged, but it is for sure that, in the absence of monetary policy, prices and interest rates must fall due to capital inflow and rise due to capital outflow.

Second, the model is used to analyse the effects of labour migration. Differences in real wages between localities accelerate migration, which affects both aggregate supply and aggregate demand conditions in out- and in-migration localities. The real effects on the supply side depend merely on the amount of migration, but the nominal effects, namely those on local prices, nominal wages and interest rates depend on the effects of migration on local aggregate demand. The demand conditions determine the particular adjustment path. The typology of *price adjustment, nominal wage adjustment, opposite adjustment* and *parallel adjustment* is again utilised in the analysis. Since the employment effects and the real wage effects of migration are now of opposite signs, the adjustment path depends on whether or not the employment effects dominate the wage effects in determining local consumption demand. If the employment effect substantially dominates the real wage effect, then prices and interest rates move in the same direction (downwards under emigration and vice versa) and nominal wages in the other. If the dominance is not substantial enough, however, prices, nominal wages and interest rate all adjust in the parallel direction (upwards under emigration and vice versa). The key issue is that local interest rates may rise or fall due to migration depending on the particular adjustment path.

From a theoretical perspective, the path of *parallel adjustment* seems quite probable, because other kinds of adjustment paths require substantial changes in aggregate demand to start them up. The issue remains an empirical one, and depends on the real wage elasticity of labour supply and labour demand in the labour market, and on the real wage difference to be adjusted. Empirically, however, it is arguable that considerable changes do occur – there is considerable evidence that prices rather fall than rise in out-migration localities, and vice versa. This is reasonable, especially from the point of view of housing and real estate prices, which constitute a considerable proportion of local consumption expenditures, and which are determined by inelastic supply conditions, at least in the shorter term.

The analysis also illustrates a possible interior solution of spatial allocation of labour in a multi-locality economy. Migration equalises the real wages in the localities, but nominal wages and price levels are not necessarily equalised. With immobile capital and no trade, great differences in local prices, nominal wages and interest rates may thus exist.

Opposing capital movements inhibit the migration effects. This is intuitive because the capital inflow increases the domestic productivity of labour and vice versa. An inflow/outflow of capital is a substitute for emigration/immigration, and it can fully offset the need for migration if the impact of capital accumulation is strong enough to make the domestic real wage the same as the market wage in the outer economy. Moreover, the impact may even be high enough to make the domestic real wage higher/lower than that of the outer economy, thus the initial migration pattern is reversed.

Finally, the analysis shows that when both labour and capital are mobile simultaneously, it is feasible that there exists an interior market solution of spatial factor allocation between localities. The result holds, however, only in the special case, where the factors are initially drawn to opposite

directions by differences in real wages and interest rates, and where the price level is fixed so that the possibility of *parallel adjustment* is ignored. Theoretically speaking, the conclusion then is that, in the present model without further restrictions and policy considerations, the market mechanism is incapable of securing a stable and efficient interior solution to factor allocation. Corner solutions, where some localities flourish and some others end up deserted, may well exist. In the real world, the question remains empirical in nature.

## References:


Andersson, Å. and R. Kuenne (1986): *Regional Economic Dynamics.* In: P. Nijkamp (ed.), Handbook of Regional and Urban Economics. Amsterdam: Elsevier Science Publishers.

Barro, R. and X. Sala-i-Martin (2004): *Economic Growth.* 2nd ed. Cambridge (Ma.): The MIT Press.

Brown, C. and P. Jackson (1978): *Public Sector Economics.* Oxford: Martin Robertson & C. Ltd.

Burda, M. and C. Wyplosz (1997): *Macroeconomics. A European Text.* Oxford: Oxford University Press.

Copeland, L. (2005): *Exchange Rates and International Finance.* Harlow: Prentice Hall.

Fujita, M. (1989): *Urban Economic Theory. Land Use and City Size.* Cambridge: Cambridge University Press.

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

Heijdra, B. and F. van der Ploeg (2002): *Foundations of Modern Macroeconomics.* Oxford: Oxford University Press.

Isard, W. (1975): *Introduction to Regional Science.* Englewood Cliffs: Prentice-Hall Inc.

Isserman, A, C. Taylor, S. Gerking and U. Schubert (1986): *Regional Labor Market Analysis.* In: P. Nijkamp (ed.), Handbook of Regional and Urban Economics. Amsterdam: Elsevier Science Publishers.

Krugman, P. and M. Obstfeld (2003): *International Economics. Theory and Policy.* 6th ed. Boston: Addison-Wesley.

Krugman, P. and R. Wells (2005): *Microeconomics.* New York: Worth Publishers.

Laakso, S. and H. Loikkanen (2004): *Kaupunkitalous. Johdatus kaupungistumiseen, kaupunkien maankäyttöön sekä yritysten ja kotitalouksien sijoittumiseen.* Helsinki: Gaudeamus.

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

Richardson, H. (1978): *Regional & Urban Economics.* Harmondsworth: Penquin Books Ltd.

Figure 1: General equilibrium properties of the model

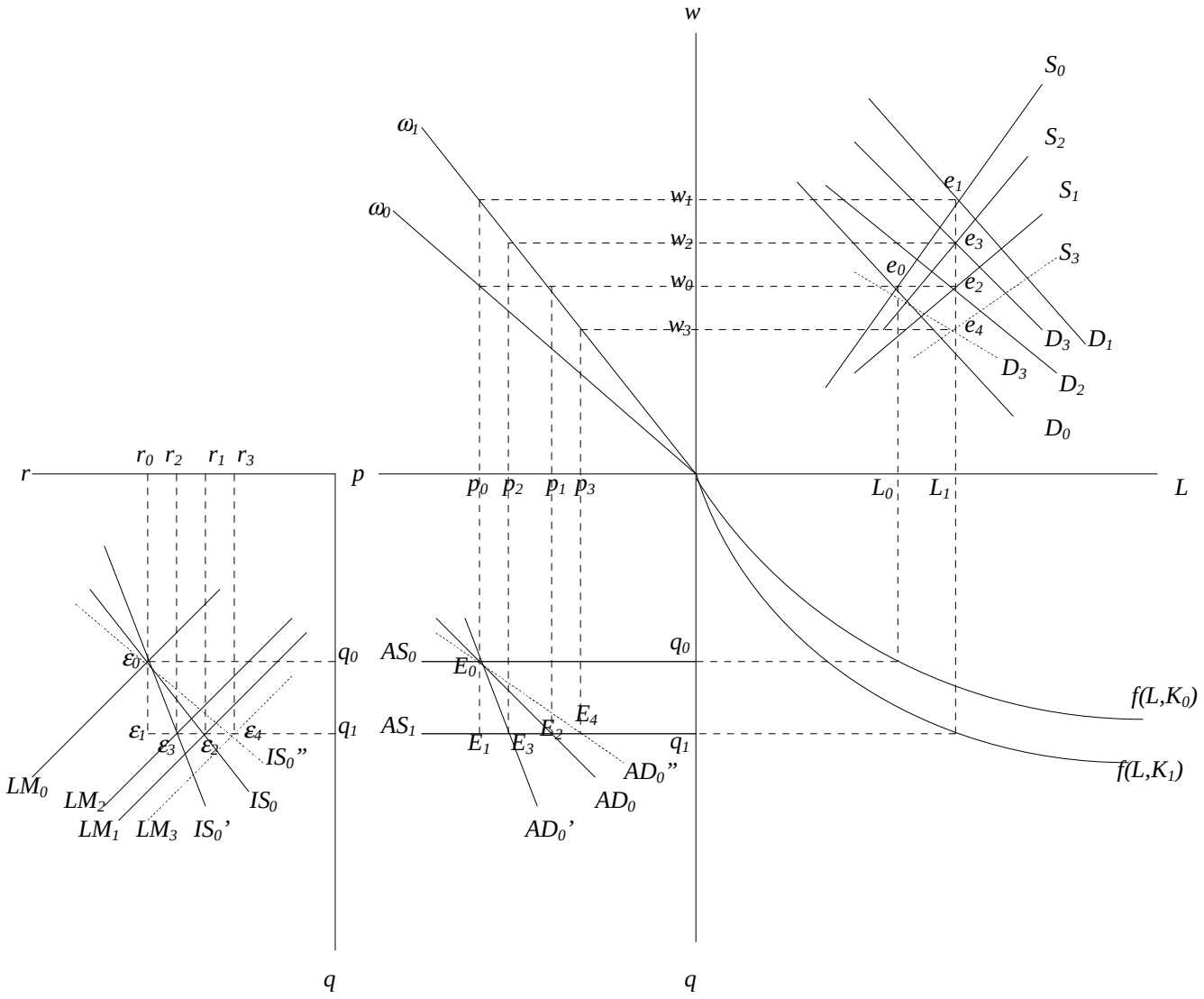


Figure 2: Emigration as a response to real wage differentials

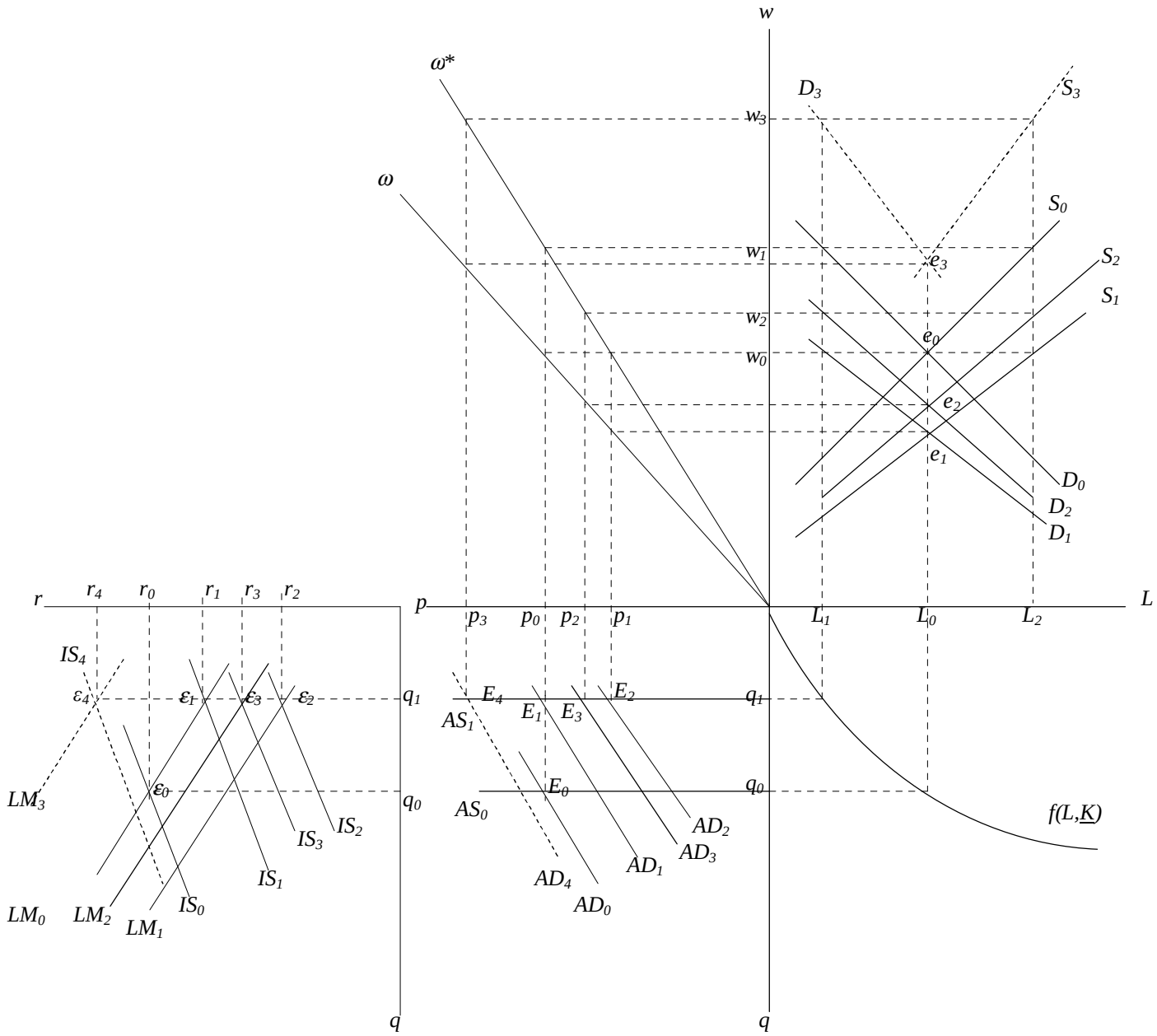


Figure 3: Migration equilibrium between two large localities

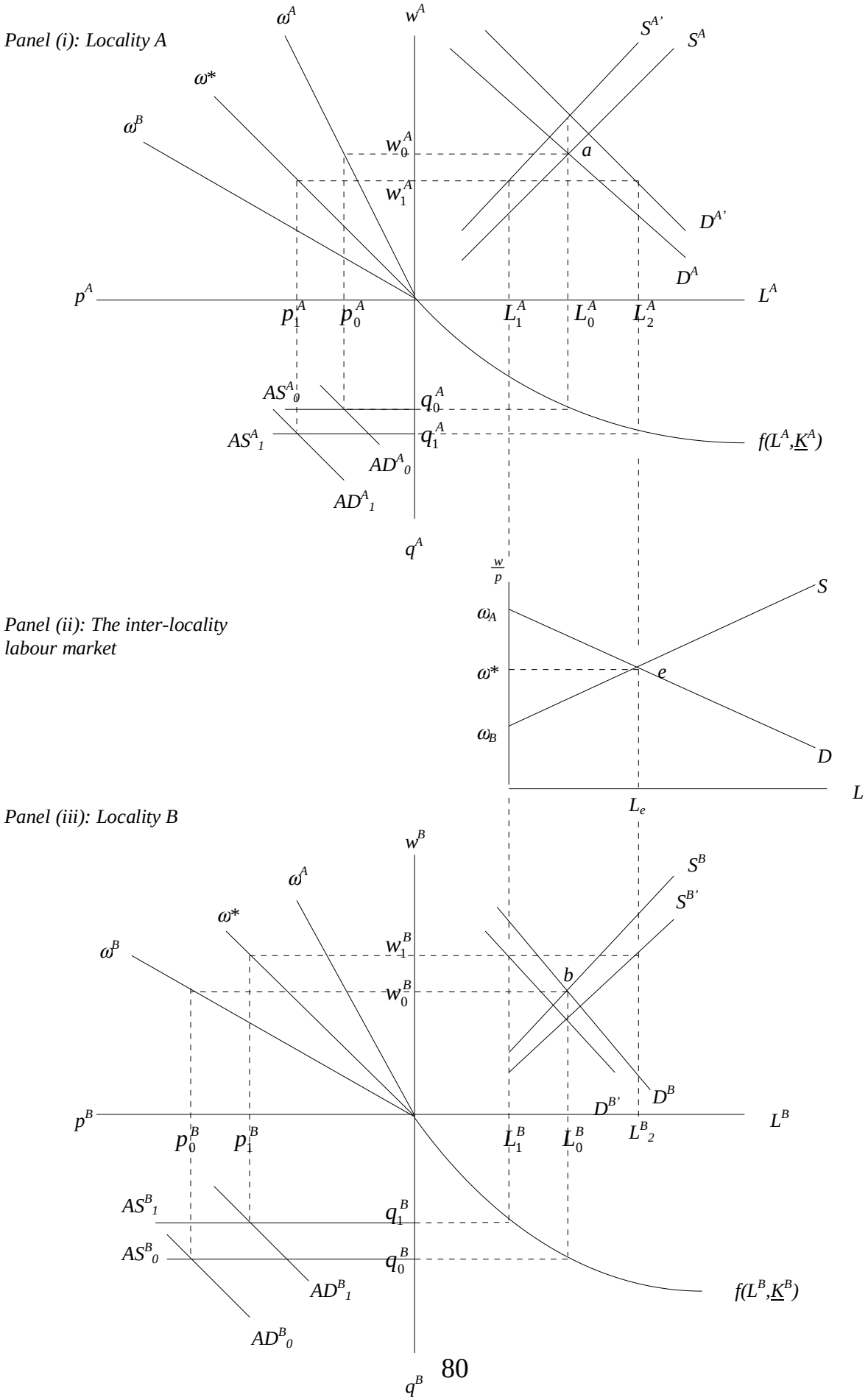


Figure 4: Capital inflow and emigration as substitutes

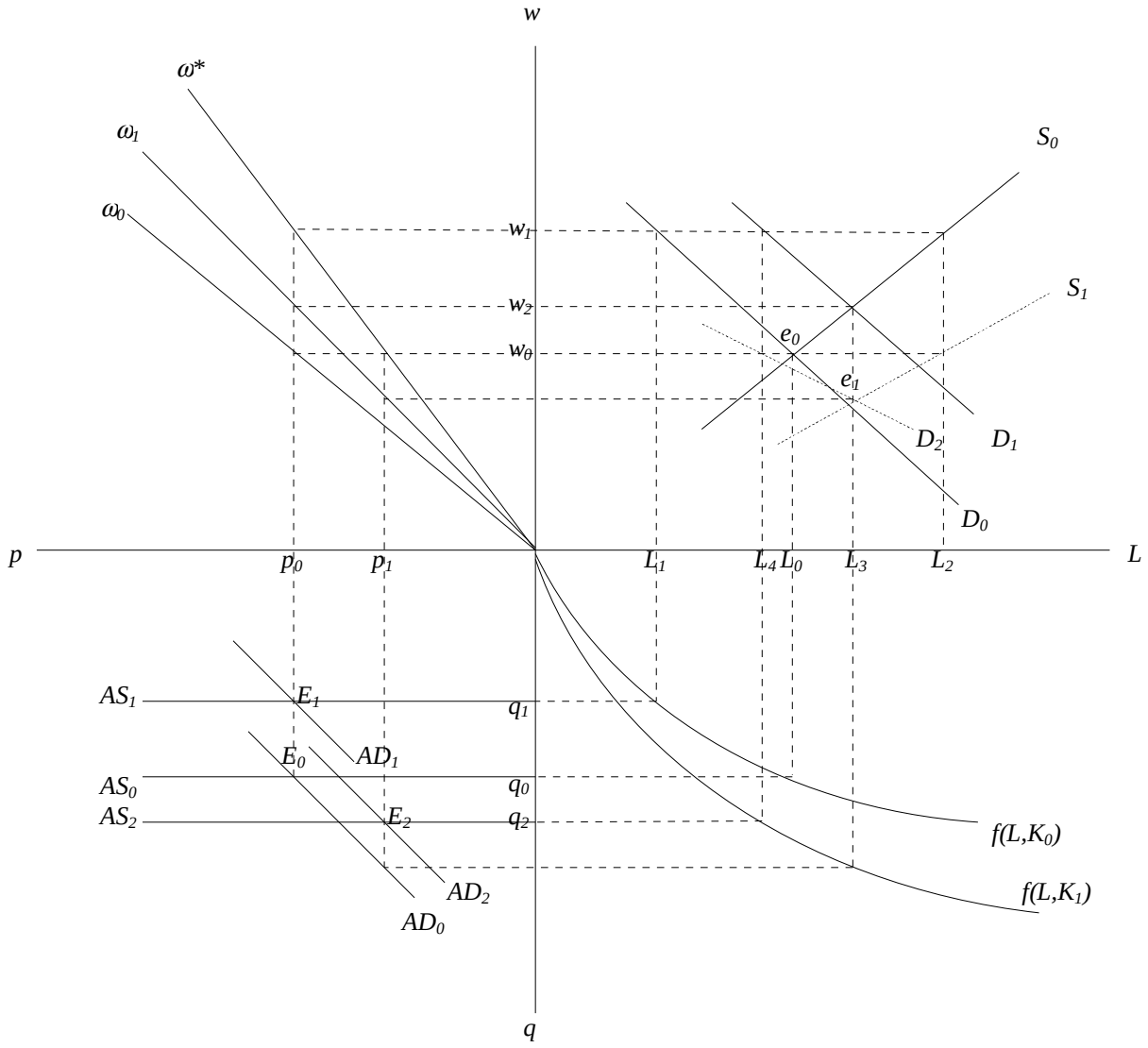
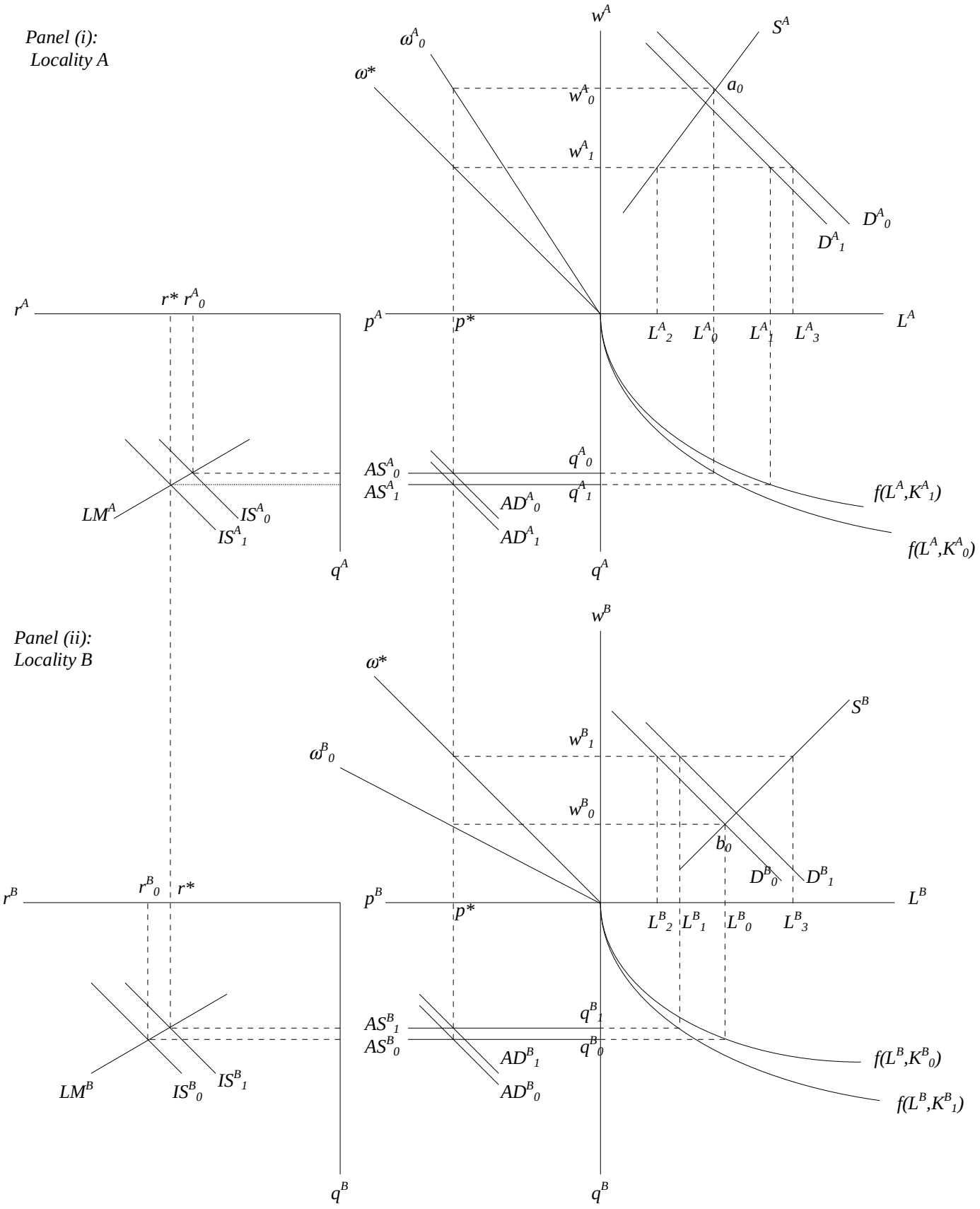


Figure 5: Capital movements and migration between two localities



# Essay #3: Taxes, Transfers and Migration

***Abstract:*** *A neo-Keynesian macroeconomic elaboration of the classic model of inter-regional migration is constructed to examine the effects of local public policy on migration. An application of the concept of the Adaptive Expectations Hypothesis is used to describe people's perceptions of the programme. It is shown that a fair programme, which does not distort the real wage comparisons between localities, does not affect migration. On the other hand, an unfair programme is shown to encourage migration the farther the anticipations concerning the repayments are from the true nature of the programme. In the longer term, when people correct their perceptions regarding the repayments, the effects are mitigated and eventually reduced to zero, provided that the programme is actually fair.*

*Key words: adaptive expectations, fair/unfair programme, labour market*
*JEL classification: 931*
*An earlier version of the paper was published under the title 'Effects of a Tax-Transfer Scheme on Migration' in: Kanniainen V.& S. Kari (eds.), Taxation, Economic Policy and the Economy, VATT-Publications, 41, 2005.*

## 1 Introduction

In the regional and urban economics literature, the classic theory of inter-regional migration is based on the analysis of the interplay of local labour markets. In the model, perfectly free and costless migration is motivated by inter-regional differences in welfare, measured in terms of real wages. Migration equalises the differences between localities and ends up with an efficient equilibrium where nobody can benefit from relocation.

The tax competition literature, on the other hand, stresses that any local policy measures which distort the welfare comparisons between localities must affect migration. Clearly, this is a constraint on inter-locality policy variations. A common view is that the pressure of migration also equalises taxation between localities – differences in local policies should be unsustainable in the longer run. The ultimate outcome would be the so-called "race-to-the-bottom", in which the public programmes are everywhere driven to their minimum levels.

Combining the two views above necessitates that both the revenue and the expenditure sides of local public policy programmes are considered simultaneously. In the spirit of the classic real wage

model of migration, a reasonable argument is then that if the taxes paid are fully compensated in terms of public provision so that the real wage remains unaffected, the free migration equilibrium should remain intact.

This paper investigates the effects of local public policy on migration in a neo-Keynesian general equilibrium elaboration of the classic labour market model. A simple programme of wage taxes and consumption price subsidies is introduced, and its effects on local market conditions and migration are studied. In deriving the main results, an application of the Adaptive Expectations Hypothesis (AEH) is utilised with regard to the anticipated repayments from the programme. The programme is supposed to be fair in the sense that redistributive policy goals as well as possible externalities caused by the programme are excluded. The results confirm the basic intuition that a fair programme does not distort migration. The main virtue of the analysis is in yielding valuable insights into what happens while the programme runs from its introduction to full maturity.

The paper proceeds as follows. Section 2 presents the basic model of the local market circumstances under autarchy in an economically determined market area, which is henceforth called a locality. A tax-transfer programme is incorporated, and the effects of fair and unfair programmes on the local market equilibrium are studied. Section 3 presents a graphical treatment of the properties of the model. The short-term and long-term emigration effects of the policy are examined in Section 4. Section 5 concludes.

## 2 The model

The model is an elaboration of the seminal neo-Keynesian macroeconomic model (for more conventional presentations see e.g. Brown & Jackson, 1978, p. 286-294; Heijdra and van der Ploeg, 2002, p. 8-12). In the model, output is given by the production function

(1) $q = f(L,K)$,

where the capital stock, $K$, is constant in the short term. Therefore, production depends on the amount of labour, $L$, measured in terms of labour time units. The standard neoclassical assumptions concerning the production function are made, namely constant returns to scale and $f_1 > 0$, $f_{11}$, $f_{22} < 0$, $f_{12} = f_{21} > 0$, where the subscripts refer to first and second derivatives of the function with respect to its arguments in order of appearance in function (1).

Define the short-run profits in the firm sector of the economy as $\pi = pq - wL$, where $p$ is the market price and $w$ is the market reward for labour time. Recalling (1), competitive profit maximisation with respect to labour use yields

(2) $w = pf_1$

for the demand for labour in the economy, written in terms of nominal wages. By function (2) the demand for labour input equals the market value of the marginal product of the input and, following from the assumption of diminishing marginal product of labour, the demand curve is unambiguously downward sloping. The nominal price level and the marginal product of labour together determine labour demand in the economy.

The household sector maximises utility subject to a budget constraint. The budget constraint is affected by a local tax-transfer programme which imposes a tax on labour income and uses the tax revenue for transfers to the consumers. The tax rate is $t$, $0<t<1$, on nominal wages. The respective transfers are given in the form of a price subsidy of rate $s$, $0<s<1$. The net nominal wage for the

workers is *(1-t)w*, and the net consumption price level is *(1-s)p*. The utility maximisation problem then reads

(3) *Max U(q,1-L) s.t. (1-t)wL =(1-s) pq.*

Solving problem (3) with respect to time use yields the following labour supply function:

(4) $w = \frac{1}{\alpha} pg(L),$

where $\alpha = (1-t)/(1-s) > 0$ is the implicit rate of return (or the repayment rate) of the programme and where *g(L)* describes the valuation of time respective to its opportunity cost. It is henceforth assumed that the opportunity cost of labour time is, on the aggregate, positive, $g' > 0$, namely the substitution effects dominate the income effects in supplying labour time. The labour supply curve is therefore by assumption upward sloping.

A fair tax-transfer programme is a scheme where the wage taxes are fully compensated to the taxpayers in terms of consumption price subsidies. That is, $t = s$ and $\alpha = 1$, thus the workers' real wage, $\omega = (1-t)w/(1-s)p$ remains unaffected. The repayment rate of the programme is one-to-one.

Under a sub-fair tax-transfer programme $t > s$ and $\alpha < 1$, implying that the repayment rate of the programme is less than one-to-one. A programme may turn out to be sub-fair for several reasons. First, the programme may be intentionally sub-fair because of redistributive aims or other such policy goals. Second, the programme may create negative externalities in the economy, which causes welfare losses to the workers. Third, the programme may end up sub-fair because of

administrative or other such transaction costs, or due to inefficiencies in the public procedures. Fourth, people may be myopic and have adaptive expectations concerning the repayment rate of the programme. This is reasoned by the gradual learning process of the actual changes in prices – people get the information step by step when they make their occasional purchases (about AEH, see Heijdra & van der Ploeg, 2002, p. 31-35). In the following, only the third and especially the fourth explanation are taken into account.

If $s > t$ and $\alpha > 1$, the repayment rate of the programme is higher than one-to-one. Such a programme can be labelled super-fair. In the present model a super-fair programme might be reasoned by positive externalities raised by the policy, or by intentionally redistributive policy goals. In the following, both of these explanations are again ruled out, and the only possible reason for super-fairness is taken to be that people may have over-optimistic expectations regarding the repayment rate of the programme. However, such an interpretation is seldom used in the original AEH presentations, because it causes some inconvenience to the stability of the goods market equilibrium.

Given that the capital stock is constant, the production function together with the labour market conditions determines equilibrium output which is the aggregate supply of goods in the economy. Under a fair tax-transfer programme with $\alpha = 1$, equations (2) and (4) state that aggregate supply is invariant to the price level. That is, the long-term aggregate supply curve of the economy is vertical in $p$-$q$ space. Exogenous changes, such as those in the capital stock, shift the vertical long-term aggregate supply curve horizontally.

The short-term properties on the supply side of the model can be seen by totally differentiating functions (2) and (4), and manipulating. Evaluating at $\alpha = 1$, $dK = 0$ yields

$$\frac{\partial L}{\partial \alpha} = -\frac{g(L)}{f_{11} - g'} > 0,$$

showing that a rise/fall in the implicit rate of return of the programme makes the labour market equilibrium shift outwards/inwards along the labour supply curve. The corresponding effect on production, after totally differentiating function (1), solving for *dL* and substituting, reads

$$\frac{\partial q}{\partial \alpha} = -f_1 \frac{g(L)}{f_{11} - g'} > 0,$$

showing that the induced change in production is the change in labour supply times the marginal product of labour. That is, for a given capital stock, the higher/lower is the implicit rate of return of the program the higher/lower is the short-term output of the economy.

The effect of a change in the repayment rate $\alpha$ on workers' real wage is obvious. To see the corresponding effect on the real wage faced by the employers, totally differentiate functions (2) and (4), substitute for *dL* from the latter to the former, manipulate and have

$$\frac{d\omega}{d\alpha} = -f_{11} \frac{g(L)}{f_{11} - g'} < 0,$$

showing that the employers face a fall/rise in the real wages because of the induced rise/fall in labour supply. By the above properties, the short-term aggregate supply curve is not necessarily vertical.

The demand side conditions of the local economy are determined by the Keynesian IS-LM model. The equilibrium of the real side of the economy, the IS curve, is given by

(5) $q = c(q^D) + i(r)$,

where $c$ denotes consumption determined by consumers' disposable income $q^D$, and $i$ denotes investment determined by the market interest rate $r$. By assumption, $c' > 0$ and $i' < 0$. The IS curve is thus declining in $q$-$r$ space. The budget constraint in (3) states that consumers' disposable income is $q^D = \alpha Lw/p$, which implies that a rise/fall in the repayment rate $\alpha$ shifts the IS curve outwards/inwards.

The corresponding monetary equilibrium of the local economy, the LM curve, is given by the money market equation

(6) $m/p = l(q,r)$,

where the left side is the real supply of money (the amount of nominal money $m$ deflated by the price level), and the right side is the demand for money. The demand for money comprises two purposes: transaction use determined by production $q$, and inter-temporal use determined by the opportunity cost of holding liquid money, the market interest rate $r$. By assumption, $l_1 > 0$ and $l_2 < 0$. The LM curve is upwards sloping in $q$-$r$ space for given real money supply. Changes in nominal money and/or in prices affect the amount of money, manifesting in outwards and inwards shifts of the LM curve.

The derivation of aggregate demand on the goods market from the IS-LM equilibrium can be carried out by totally differentiating (5) and (6), and manipulating. The expression for the AD curve then reads

$$(7)\quad \frac{dq}{q} = \frac{1}{q}\frac{l_2}{l_2 + i'l_1}\left[c'q^D\left(\frac{dL}{L} + \frac{dw}{w} - \frac{dp}{p} + \frac{d\alpha}{\alpha}\right) + \frac{i'}{l_2}\frac{m}{p}\left(\frac{dm}{m} - \frac{dp}{p}\right)\right].$$

Expression (7) states, first, that the AD curve is falling in *q-p* space. Secondly, exogenous increases in employment, real wages and real money supply shift the AD curve outwards and vice versa. Third, since $d\alpha/\alpha = ds/(1-s) - dt/(1-t)$, higher/lower taxes shift the AD curve inwards/outwards, whereas corresponding changes in price subsidies are read to the right or to the left along the relevant AD curve.

The macroeconomic equilibrium is produced by the simultaneous price adjustment mechanism on goods and labour markets. Aggregate demand on the goods market is in a focal position, because it sets the price level in the economy, and therefore determines the short-term market outcomes. It must also be noted that the aggregate supply schedules are somewhat unconventional in this model, because firms and workers have different market information due to the tax-transfer programme.

## 3 Illustration of the model

Figure 1 illustrates the macroeconomic equilibrium of the basic model. The figure consists of a four-quadrant presentation of the labour market in the northeast quadrant, the production function in the southeast quadrant, the aggregate goods market in the southwest quadrant and the real wage in the northwest quadrant. The IS-LM framework with its reference to the aggregate demand curve is presented in the lower two-quadrant setup.

(Figure 1 about here)

In Figure 1, the autarchy labour market equilibrium $e_0$ corresponds to nominal wages $w_0$ and prices $p_0$, which give $w_0/p_0 = \omega_0$ for the equilibrium real wage. Employment is $L_0$ and production is $q_0$. The aggregate supply schedule is presented by the vertical graph $AS_0$ in the southwest quadrant of the figure. The relevant aggregate demand curve $AD_0$ is assumed to set the equilibrium price level at $p_0$ so that the goods market is initially in equilibrium at point $\varepsilon_0$.

In the spirit of the AEH interpretation of the model, Figure 1 considers only fair and sub-fair tax-transfer programmes. The effects of a super-fair programme can be derived as a mirror image, and they are commented later. Start by examining the effects of the programme on the demand for labour. Supposing that the tax is levied on the firms, the wage tax $t$ imposes a tax wedge $tw$ in the labour market, and results in a split of the labour demand curve into two curves $D_0$ and $D_1$. The former is the gross wage curve encountered by the firms. Since the firms do not receive transfers, the demand side remains to be determined by the marginal physical product of labour described by $D_0$. The $D_1$ curve is the net wage curve encountered by the workers. The curve is flatter than $D_0$ because the constancy of the tax rate $t$ implies that the tax wedge $tw$ is a constant proportion of the available gross wage, given by $D_0$. Therefore, the tax wedge becomes narrower in absolute terms as $L$ increases.

The new labour market equilibrium depends on the reaction of labour supply to the programme. Consider first the case of pure taxation without any price subsidies, $t > 0$, $s = 0$ and $\alpha = 1\text{-}t$. As consumer prices remain unchanged, the tax wedge implies that labour supply is determined at $e_1$ and labour demand at $e_2$ in the labour market. As a consequence, employment falls to $L_1$ and production falls to $q_1$. The gross nominal wage rises to $w_1$ and the net nominal wage falls to $w_2$. Prices being $p_0$ for both firms and workers, the real wage rises to $\omega_1$ for the firms, and falls to $\varpi_1$ for the workers. Since $q_1$ is produced for $p_0$ prices, the aggregate supply curve turns to $AS_1$, which is

horizontal (or Keynesian) in shape. Recall, however, that the AS curves drawn in the figure are not of normal type, because they reflect the tax wedge – they present the combinations of firms' responses to gross nominal wages and unsubsidised prices, and workers' responses to net nominal wages and subsidised prices. On the demand side, taxation shifts the IS curve inwards from $IS_0$ to $IS_1$ and the aggregate demand curve from $AD_0$ to $AD_1$. The goods market equilibrium shifts consequently to $\varepsilon_1$. In the IS-LM model, the equilibrium shifts from $E_0$ to $E_1$, implying that the interest rate falls from $r_0$ to $r_1$ because of the fall in consumption demand caused by taxation.

Another benchmark case is that of a fair tax-transfer programme, $\alpha = 1$, where taxes are fully compensated in the form of price subsidies, $s = t$. Therefore, the real wage of the workers remains unaltered at $\omega_0$. While taxation causes the perceived labour demand curve to shift inwards to $D_1$, the subsidy-induced fall in prices now shifts the labour supply curve outwards to $S_1$. The new labour supply equilibrium is at $e_0'$, horizontally just below the demand equilibrium $e_0$. The two points are separated by the tax wedge. Employment and production remain respectively at $L_0$ and $q_0$. The aggregate demand curve remains at $AS_0$ in its classical regime position. The fall in the subsidised consumer prices from $p_0$ to $p_1$ causes the goods market equilibrium to slide along the aggregate demand curve $AD_1$ from $\varepsilon_1$ to $\varepsilon_2$. The reduction in consumer prices shifts the LM curve from $LM_0$ to $LM_1$. This is reasonable because the liquid money needed in purchases of goods is partly replaced by 'public' money, such as vouchers. The equilibrium is at $E_2$, and the nominal interest rate has fallen to $r_2$. The real interest rate remains unchanged.

The case of partial repayments, $0 < \alpha < 1$, quite naturally ends up at a solution in between the above two extreme cases of Figure 1. Supposing that the workers perceive a partial fall in consumer prices, say from $p_0$ to $p_2$, they are induced to work less than under a fair system, but more than without any repayments. Their labour supply curve thus settles at $S_2$, and the labour supply

equilibrium $e_3$ implies that the real wage is $\varpi_2$ for workers and $\omega_2$ for firms. As a result, employment is $L_2$, and production is $q_2$. Viewed from the supply side, the $q_2$-$p_2$ combination shows that the aggregate supply curve $AS_2$ has turned upward sloping. The corresponding goods market equilibrium is at $\varepsilon_3$, at the intersection of the $AS_2$ and $AD_1$ curves. The IS-LM equilibrium is at $E_3$, and the equilibrium interest rate is $r_3$.

Figure 1 does not present the case of a super-fair programme, but an interested reader can conduct the experiment by assuming that consumption prices are anticipated to be even lower than $p_1$, which shifts the labour supply curve outwards beyond $S_1$. Short-term employment and production increase respectively from $L_0$ and $q_0$, and the short-term aggregate supply curve becomes downward sloping. The goods market equilibrium is found along $AD_1$ inwards from $\varepsilon_2$.

Under the AEH interpretation of the model, the workers gradually adapt towards a fair programme. Provided that the tax-transfer programme is actually fair, that is, supposing that the tax revenue is not used to promote other policy goals or to cover transaction costs and inefficiencies, the workers continuously receive new information concerning the actual (subsidised) prices and are therefore justified in changing their perceptions concerning the true repayment rate of the programme. The above two solutions regarding the sub-fair case (the cases of horizontal and upward sloping aggregate supply curves), as well as that of the super-fair case (downward sloping aggregate supply) are thus short-term in nature. In the long term, labour supply and production converge respectively towards $L_0$ and $q_0$. The long-term aggregate supply curve eventually converges with the vertical $AS_0$ schedule, leaving everything unchanged in real terms.

## 4 Effects on migration

For a simple introduction of migration to the model, assume that the local capital stock is fixed and immobile and that there is no trade between localities, but allow for labour migration in response to inter-locality differences in real wages. Assume that people are perfectly able to monitor the real wage differentials, and that migration is costless. Assume also that the locality considered is of atomistic size so that migration does not change the circumstances in the competitive total-economy labour market.

Suppose that the locality considered is initially in its long-term market equilibrium, which also matches the total-economy equilibrium. That is, the equilibrium real wage is assumed initially to be equal to that in the rest of the economy, $w_0/p_0 = \omega_0^*$. In this situation, introduce a local tax-transfer programme and examine its effects on migration. Figure 2 provides the analysis.

(Figure 2 about here)

In Figure 2, the initial market equilibrium in the locality considered is constituted by the labour market equilibrium $e_0$, which yields employment $L_0$ and production $q_0$. Nominal wages $w_0$ and prices $p_0$ give $\omega_0^*$ for the real wage. The goods market equilibrium is at $\varepsilon_0$ at the intersection of $AS_0$ and $AD_0$ curves. Next, introduce the policy and analyse its effects from the point of view of the AEH interpretation of the model.

In the very short term (the case of pure taxation) the policy splits the net-of-tax labour demand curve from $D_0$ to $D_1$ in the north-east quadrant of Figure 2. In autarchy, the local labour market equilibrium would shift to $e_1$ at a lower real wage $\varpi_1$. However, since the workers can be employed elsewhere for the constant real wage $\omega_0^*$, the result is emigration, which amounts to $L_0 - L_1$ in terms

of labour time units. Since local employment falls to $L_1$, local production also falls to $q_1$. Because of the exogenous shock in employment, the vertical aggregate supply curve shifts to $AS_1$ and the aggregate demand curve shifts to $AD_1$ so that the goods market equilibrium is at $\varepsilon_1$ for the unchanged consumer price level $p_0$. Note that while the local nominal wage for workers remains at $w_0$, that paid by the employers rises to $w_0$', and their real wage also rises to $\omega_1$. This is what they have to pay to recruit workers – the high gross wage does not attract immigrants, who would also be subject to taxation.

In the longer term (the case of partial repayments) the workers gradually start to reconcile the price subsidies, and increase their labour supply in response to the perceived improvement in real wages. When the anticipated consumer price level falls to, say, $p_2$, workers' real wage rises to $\varpi_2$, and the labour supply curve shifts to $S_1$. Reading the adjustment at the price level $p_2$ from $\varpi_2$ to $\omega_0$*, emigration now amounts to $L_3 - L_2$. The goods market equilibrium shifts to $\varepsilon_2$ on $AD_2$. The key finding is that the amount of emigration resulting is unequivocally smaller in this case than in the above case of pure taxation.

In the long term (the case of a fair system) the price subsidies are fully anticipated, and consumer prices fall from $p_0$ to $p_2$. The initial real wage $\omega_0$* is restored. The goods market equilibrium is at $\varepsilon_3$ and no emigration emerges. The programme results in a tax wedge in the labour market and in a corresponding wedge in prices on the goods market, which cause the aggregate demand curve to shift to $AD_3$ (which corresponds to the $AD_1$ schedule in Figure 1). The conclusion is that a fair tax-transfer programme does not distort migration.

Figure 2 does not present the case of a super-fair programme, but an interested reader can again conduct the experiment with prices lower than $p_2$, and shifting the labour supply curve beyond $S_2$.

Since the workers' real wage now rises above $\omega_0$*, the result for migration is the very opposite to that derived in the sub-fair case – in the short term, people are induced to immigrate into the locality considered. Employment and production rise, and the AS and AD curves shift outwards.

The above experiments follow the basic intuition. It is obvious that taxation that cuts real wages encourages emigration. It is also obvious that a fair programme has no effects on migration. The most important message of the analysis, however, is in between these two extreme cases. Since this is also the inter-space in which most of the policy programmes operate in practice, it is useful to understand the orientation of the equilibrium path. The AEH interpretation of the model provides an important description of orientation. By this interpretation, people gradually adapt to the fair nature of the programme. New programmes may markedly increase migration, but the effect is eventually mitigated, and the long-term effect converges to the zero effect of a fair programme. Another consideration is that administrative costs, inefficient conduct of the program etc. may easily press the repayment rate below one-to-one so that the concept of a fair program is quite idealistic in practice. In any case, there is good reason to believe that the long-term policy effects on migration are quite insignificant.

## 5 Conclusions

The paper provides a simple and illustrative model for the analysis of the effects of local public policy on inter-locality migration. A neo-Keynesian macroeconomic model is used, and an Adaptive Expectations Hypothesis type description of people's perceptions of the policy programme is utilised to explain the short-term deviations between perceived and true market conditions. The framework yields a comprehensive clarification of a complicated issue concerning the effects of local policy on inter-locality migration.

Literally, a local tax-transfer programme of wage taxes and consumption price subsidies is studied in the model, but the message of the analysis can be generalised to concern the provision of local public goods and services. The tax wedge in the local labour market can be interpreted as the share of public production, and the corresponding price wedge on the local goods market can be interpreted as the share of free public goods, or transfers-in-kind, in the local economy.

The effects of the policy programme have been shown to depend on whether the scheme is fair or unfair, which again depends on if taxes are expected to be fully compensated in terms of monetary or real repayments or not. Even an actually fair programme may be anticipated to be sub-fair in the short term, during the AEH type adaptation path towards the long-term equilibrium.

The main lesson from the analysis is four-fold. First, a fair and correctly perceived policy programme does not affect migration. This is because a fair system leaves the local real wage unchanged and thus has no effects on inter-locality welfare comparisons. The free migration equilibrium remains efficient. The result holds in a perfect foresight type world, around a market solution that is long-term in nature.

Second, sub/super-fair policy programmes have short-term effects that encourage emigration/immigration. The effect is the stronger the farther away the short-term equilibrium is from the long-term equilibrium. As to the effects of a sub-fair programme, a pure tax system with no (anticipated) repayments quite intuitively has the strongest effect on emigration, but even partial(ly anticipated) repayments from the system undermine the motivation for emigration. The short-term migration equilibrium is distorted because the real wage faced by the workers deviates from that faced by the firms.

Third, the short-term effects of sub/super-fair programme are eventually dampened in the longer term, provided that the tax revenue is not wasted, there are no redistributive policy goals, and/or there are neither positive nor negative externalities involved in the programme. In these circumstances an AEH type adjustment in the supply of labour leads to convergence towards an efficient long-term equilibrium.

Finally, the analysis also describes what happens to the local economy when part of it is 'socialised', that is if a part of people's consumption needs is provided publicly free of charge through the tax payment. The result is that, in the conditions specified above, the economy remains unchanged in real terms. In other words, there may well be policy differences between localities, at least as far as there are no inter-locality spill-overs. Free migration thus does not automatically lead to equalisation of policy programs between localities, and tax competition does not necessarily accelerate a race-to-the-bottom in terms of public spending.

The analysis is somewhat simplistic in the sense that redistributive and other such policy goals are ignored. In practice, most policy programmes usually have effects of this type. Likewise, the analysis ignores any kinds of externalities, positive or negative, involved in the programme. It is arguable that practical policy programmes are often justified by the promotion of positive externalities, or by the correction of negative ones. However, the main lines of the results can also be read from these points of view – the unequivocal migration effects and market distortions of intentionally unfair programmes must be accounted for either as targets or as opportunity costs of the policy. From this perspective, the question of tax competition and sustainability of the policy also enters into the picture. For a proper consideration of positive and/or negative externalities caused by the programme, its effects on local real wages should be modelled nonlinearly.

## References:


Brown, C. and P. Jackson (1978): *Public Sector Economics.* Oxford: Martin Robertson & C. Ltd.

Burda, M. and C. Wyplosz (1997): *Macroeconomics. A European Text.* Oxford: Oxford University Press.

Fujita, M. (1989): *Urban Economic Theory. Land Use and City Size.* Cambridge: Cambridge University Press.

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

Hansen, J. and J. Nielsen (1997): *An Economic Analysis of the EU.* London: McGraw-Hill.

Heijdra, B. and F. van der Ploeg (2002): *Foundations of Modern Macroeconomics.* Oxford: Oxford University Press.

Laakso, S. and H. Loikkanen (2004): *Kaupunkitalous. Johdatus kaupungistumiseen, kaupunkien maankäyttöön sekä yritysten ja kotitalouksien sijoittumiseen.* Helsinki: Gaudeamus.

Laurila, H. (2004): General Equilibrium Elaboration of the Classic Model of Migration. *Tampere Economic Working Papers, Net Series,* No. 35, December 2004.

Wilson, J. (1999): Theories of Tax Competition. *National Tax Journal,* 52, no.2, 269-305.

Figure 1: Short-term and long-term effects of a tax-transfer programme

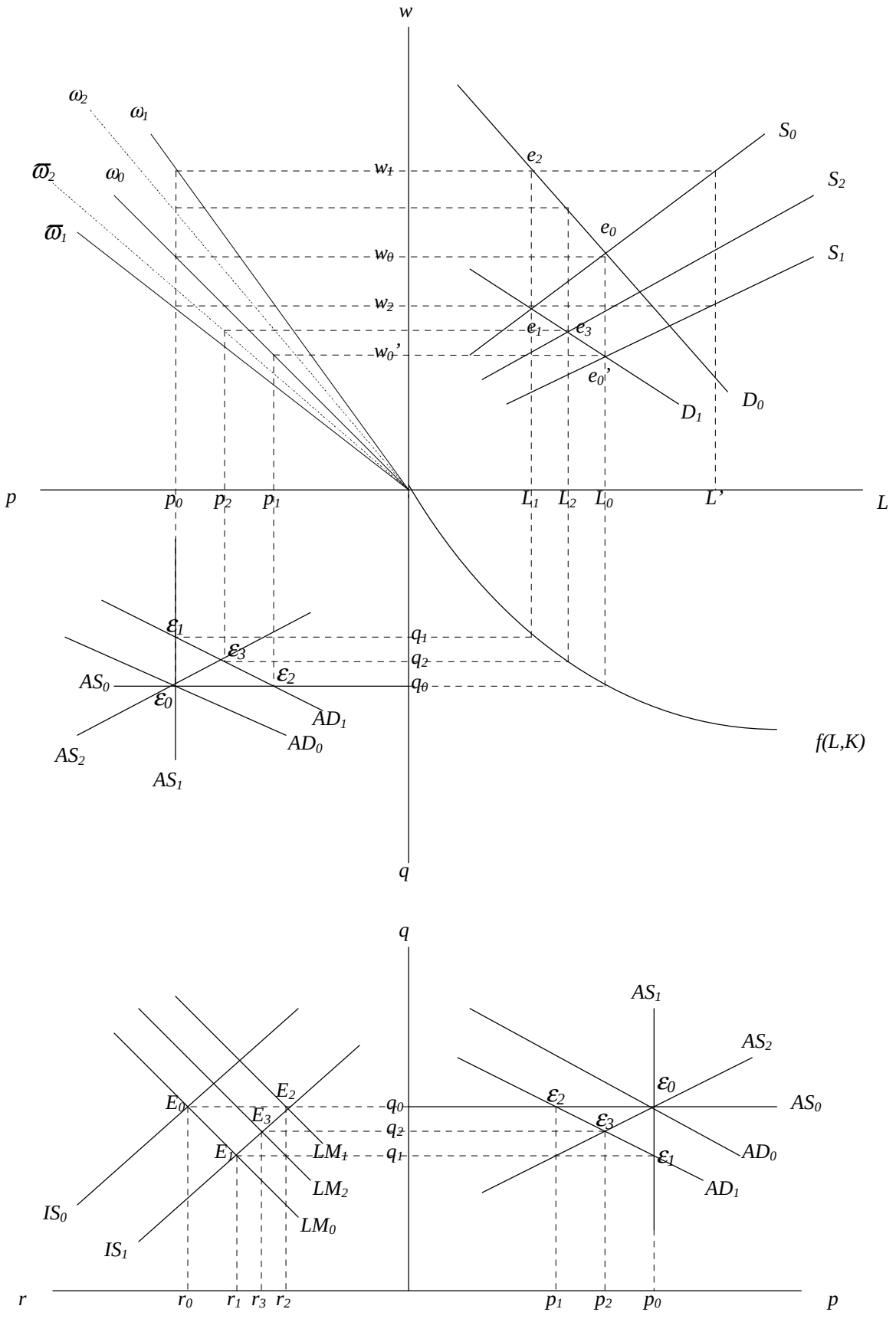
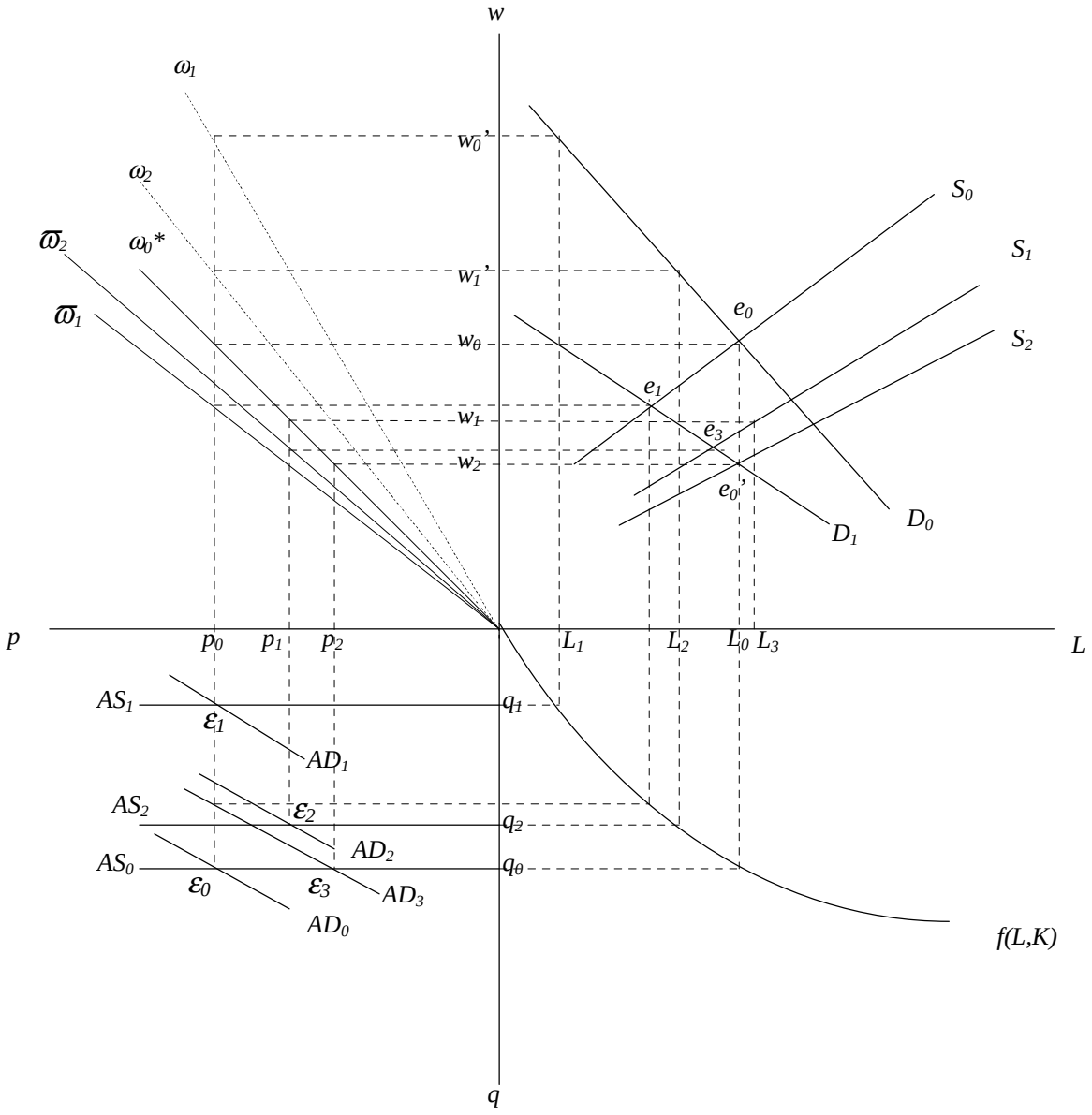


Figure 2: Effects of a tax-transfer programme on emigration



# 4. APPROACH III: THE CLUB THEORETIC APPROACH

## Essay#4: The City as a Club

***Abstract:*** *The paper presents a club theoretic model of a city. In the model welfare in a city depends on its size because of positive and negative externalities (or agglomeration economies and diseconomies) generated by the close proximity of people and economic activities. Technological externalities appear in people's utility functions and pecuniary externalities are due to shortcomings of the price mechanism. Both types of externalities are internalised in the city and act as centripetal and centrifugal forces that explain migration and thus the formation and development of the city. The issue of optimal city size is studied and it is shown that the individually originated formation mechanism of the city may lead to non-optimal solutions, even when the city itself acts as a true market agent in determining its optimal size.*

*Key words: agglomeration, city size, club theory, externalities*
*JEL classification: 931*

### 1 Introduction

Cites are spatial concentrations of people and economic activities. The dominant reasons behind the existence of modern cities are connected to the performance of the market. Since most market activities are very spatial in nature, the proximity of people and the functions that they engage with in their everyday lives has obvious implications for the welfare that the economy is able to create. People choose their place of living so as to maximise their welfare observing benefits and costs attached to different sites. The growth of some cities is simply due to their ability to create better welfare for their citizenry than those places that people are leaving. The existence and development of cities can therefore be explained by their ability to create welfare.

An economic explanation for the existence and development of cities consists of centripetal forces such that draw people closer to each other, and of centrifugal forces such that throw them apart. A standard assumption in the literature of urban economics is that there is a connection between city size and welfare. The connection can be explained by agglomeration economies that arise in the production and consumption of the outputs of the local private and public sectors. According to this

theory an increase of city size yields benefits and cost savings, but eventually it may also cause falling benefits and increasing costs. Gains from agglomeration economies are experienced as higher labour incomes resulting from specialisation and increased choices in work, wider consumption facilities of private and public goods and services, cost savings in arranging the provision of collective goods, larger variety of leisure activities and better opportunities for all kinds of interaction. Among the main disadvantages caused by agglomeration diseconomies are increased costs of housing and commuting, rising local taxes and tariffs, increased health risks and risks of property damage caused by pollution and crime, exacerbating social problems and other such congestion costs.

Agglomeration economies and diseconomies are due to externalities and economies of scale. Without such market failures a totally dispersed spatial pattern would prevail. In the presence of externalities, both positive and negative, people not only respond to the observed benefits and costs but also create them and thereby affect the welfare anticipated by others. From a club theoretic viewpoint, the essential nature of the city is to pool the positive and negative effects of agglomeration and to internalise them into people's everyday welfare experiences. Because the externality problem is so apparent in the context of a city, city size is a major economic variable - it is the optimal size of the city that facilitates full internalisation of externalities.

The paper examines the formation and development of a city based on the decision-making of utility maximising people or households. Firms are ignored, assuming that they decide on their production and location so as to maximise their profits, and the provision of local collective goods is assumed to be efficient. Thus, the decisions of firms and the local public sector, which are also affected by agglomeration economies and diseconomies, are fully reflected by the market information encountered by the people. The paper proceeds so that Section 2 discusses the connection of agglomeration economies and diseconomies and city size and Section 3 transforms individual utility

maximisation into a club theoretic framework. Section 4 illustrates the model, and Section 5 compares possible solutions for city size from the point of view of efficiency. Section 6 concludes.

## 2 Agglomeration economies and externalities

The centripetal and centrifugal forces that explain the formation of a city are here simply called agglomeration economies and diseconomies (Krugman, 1991; Fujita & Thisse, 2002). People gather together because of agglomeration economies, and they drift apart because of diseconomies that arise in overly concentrated surroundings. From the point of view of an individual household, the effects of agglomeration can be divided into direct and indirect effects (cf. Richardson, 1973).

The direct effects are the household effects experienced on the local goods and factor markets. The variety, quality and attainability of private and collective goods, occupation, housing, leisure activities, environment etc. are all locally determined and depend on agglomeration. Positive effects arise as concentration of economic activities results in greater variety of choice of private and collective goods and services, facilitates greater specialisation and remunerative occupation, and makes it possible to enjoy a wider variety of urban amusements during leisure. Negative effects arise as concentration eventually causes physical and mental congestion that impairs the enjoyment. Agglomeration diseconomies emerge in large-scale cities e.g. in the form of overcrowding, pollution and crime.

The indirect effects arise because spatial concentration of economic activities creates cost savings on the level of the local marketplace. The performance of the local economy involves a wide variety of physical contacts, travel and transport on the commodity and factor markets. All these actions are costly in terms of both money and time, and the rate of spatial concentration has a considerable effect

on the costs. The indirect effects are called business and social agglomeration economies and diseconomies.

Business firms derive several gains from the size of local factor and product markets. They may derive economies of scale in their own production decisions, localisation economies within industries sharing the same spatial marketplace, and urbanisation economies from the overall economic activity in one area. The social agglomeration economies and diseconomies are channelled through the local public sector. The local public sector can also utilise scale economies, cost sharing etc. in its service production. However, in an overcrowded city the advantages of size begin to diminish. Agglomeration diseconomies are mainly due to the fact that the city involves land use. The more agglomeration there is the scarcer land becomes. Constructing and maintaining an economically, socially and environmentally sustainable technical and social city infrastructure becomes more costly the more crowded the area is. Agglomeration diseconomies arise when congestion starts to cause extra production costs due to rising land rents and delivery costs. The business and social agglomeration economies and diseconomies are indirect in that, through the working of the local market, they eventually end up affecting the households' welfare.

Agglomeration economies and diseconomies are essentially externalities in nature in that the choices made by individual agents affect other agents' welfare in the city. Externalities are of two kinds, technological externalities that affect the utility function, and pecuniary externalities that operate through the price mechanism (Scitovsky, 1954; Fujita & Thisse, 2002).

Technological externality is the effect of an economic activity on the consumption set and the utility function of an individual. The effect concerns an agent other than the one exerting this economic activity. Technological externalities are anticipated in everyday life and they include both physical

and mental factors. Due to technical externalities the arguments in people's welfare functions differ in cities of different sizes. Technological externalities bear a close resemblance to the above definition of household agglomeration economies and diseconomies. However, business and social agglomeration economies and diseconomies affecting the consumption sets and not working through the price system may also be regarded as technological externalities.

Pecuniary externalities affect people's welfare through the price system. Those business and social agglomeration economies and diseconomies affecting wages, the prices of private goods and housing, and local taxes are such externalities. It is assumed here that the local business sector and the local public sector are competitive and effective in their decision-making, finance and production. This assumption implies that the pecuniary business and social agglomeration economies and diseconomies revert to the unbiased market price parameters encountered by the households. What is not ruled out, however, is the effects of the migrants themselves on wages, prices and taxes in the city. Migrants with market power cause pecuniary effects which they do not take into account in their own decisions. Therefore, the set of market prices may still be biased.

In the presence of externalities, individual decisions imply market power and introduce a certain collective, or club theoretic, element into the mechanism of the formation of a city (Tiebout, 1956; Buchanan, 1965). The key thing is that, from the club theoretic viewpoint, the external effects are internalised into the households' utility within a city (Cornes & Sandler, 1986; Scotchmer,1994). It must be noted, however, that there may also exist interregional externalities that are not internalised. Agglomeration boosts the systemic functions of big cities as a diffusion source for innovation and development impulses within regions and down the national urban hierarchy.

## 3 The club theoretic nature of a city

In a market economy, individuals choose their locations of residence in order to maximise utility. A standard problem of constrained utility maximisation of an individual or a household reads

(1) $Max\ U(q_x,q_y,l)\ \ s.t.\ \ w(1-l) + k = p_xq_x + p_yq_y$.

In the utility function, $q_x$ and $q_y$ are the quantities of private and collective consumption, respectively, and $l$ is leisure. These are the endogenous variables. In the budget constraint, $w$ is the market rent (or wage) for work time $1-l$, where the total available time is normalised to unity, $k$ is the non-labour income, and $p_x$ and $p_y$ are respectively the market prices for private and collective goods. These are the exogenous market parameters. By (1), the individual is assumed to derive utility from the amounts of private and collective consumption and from leisure. The standard properties concerning the shape of the utility function are assumed. According to the budget constraint, all labour and non-labour income on the income side is spent on private and collective consumption during the time period considered. Problem (1) cannot be solved without further restrictions, but the formulation suffices to illustrate the main idea of the analysis.

By the argumentation in Section 2, both the endogenous variables and the exogenous parameters in (1) are affected by externalities that are attached to agglomeration. Therefore, they can be expressed as functions of city size. Actually, it is the spatial proximity (or density) of the elements of the local market that explains agglomeration economies. However, it is simpler to use population as a measure of city size. This is reasonable, when the geographical area of the city is taken as given in the short term.

Denote the population in the city by $n$, and write $q_x = q_x(n)$ and $q_y = q_y(n)$. Thus, the available sets of private and collective goods may differ between cities of different size. Moreover, $l = l(n)$ shows that the amount and quality of leisure may vary between cities. This is explained by time saving due to specialisation in work, and by quality of life arguments. Implicitly it also includes the non-monetary emoluments of work. The connection of the endogenous variables to city size is explained by technical externalities. In the budget constraint, $w = w(n)$ refers to the wage effect of specialisation, and $p_x = p_x(n)$, $p_y = p_y(n)$ refer to the business and social effects of agglomeration on the market price of the private good and on the tax price of the collective good. Pecuniary externalities affect the formation of both prices. The non-labour income $k = k(n)$ may also be assumed to be locality-dependent insofar as the required market transactions are spatial in nature.

The club theoretic property of the city means that the agglomeration economies and diseconomies are transformed into locally experienced welfare. Now, since all the relevant arguments in the utility function and all the variables and parameters in the budget constraint depend on $n$, we can rewrite problem (1) in a standard club theoretic form (Buchanan, 1965; Cornes & Sandler, 1986) as

(2) $Max\ W\ s.t.\ W = B(n) - C(n)$,

where $W$ is individual welfare (or net benefit), $B$ is the individual benefits experienced in the city and $C$ is the individual net monetary costs of living in the city. The benefit side refers to the utility function in (1), and the cost side refers to the budget constraint in (1). Technological externalities operate on the benefit side, and pecuniary externalities operate on the cost side.

By (2), the benefits and the costs and therefore also the welfare experienced in the city depend on city size. A conventional assumption is that agglomeration economies dominate at earlier stages of

city growth, but that agglomeration diseconomies eventually start to dominate in overcrowded surroundings. On the benefit side it is quite obvious that the variety and quality of private and collective goods and leisure increase with city size, but that congestion and crowding in the streets and shops, pollution, damage, crime and other unpleasant features eventually arise and cause the benefit to fall. On the cost side, city growth promotes higher wages for less work, lowers prices of private goods and housing, cuts taxes and search costs and so on. However, as the geographical area of the city becomes excessively congested, land rents, traffic and other transaction costs, infrastructure costs etc. start to rise.

Expression (2) is written in terms of an individual household. In principle, the problem is now in solvable form since there is only one variable, namely *n*, to be solved. It cannot, however, be solved by any individual household alone, because individuals cannot choose such a variable as n, the size of a city. What they can do is to decide on whether or not to move to a city of a certain size. As an optimisation problem, expression (2) can be solved only collectively - the city as a club is able to do so provided that it has the necessary instruments. Therefore, allowing for free mobility, expression (2) has two interpretations. On the one hand, expression (2) describes how people make their location decisions by comparing all kinds of benefits and costs attached to different locations. And on the other hand, expression (2) describes the utility maximisation of an individual as a full member of the club in which he resides.

## 4 Illustration of the model

Assume that the physical size of the city is set at its long-term optimum. Assume also that there are no inter-city externalities. This is to say people's perceptions about the benefits and costs associated

with a city fully reflect social values. Provided that firms maximise profits and that the local public sector is efficient, agglomeration economies and diseconomies end up being experienced by the people in the city.

In order to fully understand the club theoretic nature of a city, start by examining the total welfare which the city can produce for its entire population (Richardson, 1973; Fujita, 1989). Define $TW = nW = nB(n) - nC(n) = TB - TC$. The total benefit function $TB = nB(n)$ is assumed to be S-shaped. Taking the geographical area of the locality as fixed in the long-term optimum, the *TB* function starts from the origin. On low population levels total benefits grow at an increasing rate. This is because increasing population generates a variety of agglomeration economies in goods and factor markets, including all kinds of monetary and non-monetary emoluments of urban life. However, the inevitable increase in congestion causes agglomeration diseconomies to arise. The rate of growth of the total benefits begins to fall at the point, at which agglomeration diseconomies gain ascendancy. The total cost function $TC = nC(n)$ is assumed to be an inverted S-shape. This is reasonable because considerable agglomeration economies are due to appear when *n* rises from very low levels. The cost function rises at a decelerating rate as agglomeration diseconomies eventually occur with increasing density of population. Agglomeration economies dominate until the fixed factor, geography, becomes scarce enough so that agglomeration diseconomies start to dominate. At this point, the total cost function starts to rise at an increasing rate.

The curvatures of the *TB* and *TC* functions are familiar from standard cost theory. In the theory of the firm, the fixed cost determines the shape of the total cost function in the short run. Here the fixed factor is the geographical area of the city. In the long run, the fixed cost is also a choice variable in the firm's decision-making, as is also the geographical area in the case of cities; in the long run, both population and the geography are chosen optimally so as achieve the highest possible welfare. At the

long-term optimum agglomeration economies are fully utilised from the household, business and social perspectives. Figure 1 illustrates the above assumptions about the total benefit and total cost functions.

(Figure 1 here)

Figure 1 is drawn on the assumption that on very low population levels the total costs of a city-like structure exceed the total benefits of utilising it, but after a certain population level, namely $n^o$, benefits exceed costs. The benefit side expanding at a diminishing rate, and the cost side eventually exploding, there will emerge another intersection point at *p*. When population in the city exceeds $n^p$, total costs will again exceed total benefits. The curve *TW*, drawn as a vertical sum of the *TB* and *TC* curves, presents the total welfare. Since the *TB* and *TC* functions include all kinds of benefits and costs perceived by the people, then *TW*, the net of *TB* and *TC* presents the attainable welfare that the locality can generate for its residents. Therefore, *TW* is the relevant measure for social welfare within the city. Without interregional externalities, the *TW* function gives the relevant measure also considered from the perspective of the whole society consisting of a multiple of localities.

Total welfare in a city is an abstract concept, and it is by no means the driving force of any market-like mechanism. In their decision-making on residential location, individuals do not care about the total benefits and costs. Instead, individuals seek for their own good, and base their migration decisions on personally perceived benefits and costs. Therefore, the concepts of total welfare presented in the upper panel of Figure 1 are not relevant for the individuals and for the working of the market mechanism. The concepts must be elaborated to match to individual perceptions.

The simplest way to proceed is to assume that all individuals are alike and derive the average per capita benefit and cost, and the respective average welfare schedules from the *TB*, *TC* and *TW* curves. These concepts are presented in the lower panel of Figure 1. Note that the scale of the vertical axis is changed from the upper panel. The per capita (or average) benefit and cost functions read $B = TB/n$ and $C = TC/n$ respectively. By the properties of the *TB* and *TC* functions, *B* is inverted U shaped, and *C* is U shaped. The marginal benefit function, $MB = dTB/dn$, is a linear approximation of the curvature of the total benefit function TB. Intuitively, *MB* is the effect that one immigrant has on the total benefits perceived by all members of the locality. By the properties of the total benefit function, the marginal benefit function *MB* is inverted U-shaped. Likewise, the U-shaped marginal cost function, $MC = dC/dn$, measures one immigrant's effect on total costs perceived by the whole population in the locality. Since *C* presents average numbers, adding a number smaller/higher than the average causes the average to fall/rise. Therefore, because *MC* describes the added numbers, it must dissect *C*, the average number, from below at the lowest point of *C*. By the same token, *MB* must dissect *B* from above at the highest point of *B*.

The average concepts illustrate the individual experience, and match the maximisation problem (2). The concepts deserve attention in two respects. First, they facilitate an accurate or even cardinal treatment of individual welfare (Ng, 2000), and second, they are straightforwardly operational in the preference-revealing mechanism of migration. The assumption of homogeneous people may sound critical, however. An alternative and less stringent interpretation might be that the average concepts are those of the representative household, and that they are eventually revealed by the systematic migration pattern. The decisions of the non-representative households thus belong to the purely stochastic element of migration.

Figure 1 is drawn on the assumption that agglomeration diseconomies appear on the benefit side later than on the cost side. The inverted U-shaped *W* curve presents the welfare per person, drawn as a vertical sum of the *B* and *C* curves. The per capita welfare is positive between $n^o$ and $n^p$, creating a motive for possible immigration from places with lower welfare. The *MW* curve presents marginal welfare and, as a presentation of marginal net benefit, it is a vertical sum of the *MB* and *MC* curves. Again by the properties of the benefit and cost functions, *MW* dissects *W* from above at the maximum point of *W*.

The graphs in Figure 1 must be considered from an ex post perspective. The functions present the benefits and costs accruing to individuals and to the city as a whole only at a certain population level. That is, the curves can actually be drawn only up to the population level to which the private decision-making has driven it. Expectations regarding their development after that point are totally irrelevant to private decisions. Particularly, a potential newcomer encounters the situation resulting from the choices made before him, and he does not take into account the effect that he may have on the situation faced by the next immigrant after him.

## 5 Three solutions for city size

According to Figure 1, the solution for optimal city size seems to be straightforward. Given the above assumptions about the forms of benefit and cost functions, and that the net of them measures social welfare, the question about the most preferred size of the city becomes a standard club theoretic optimisation problem. From the point of view of economic efficiency population should be chosen so that total welfare is maximised. In Figure 1 total welfare reaches its maximum at the highest point of the *TW* curve. The optimum condition is $dTW/dn = 0$, which yields $dTB/dn = dTC/dn$ in terms of the upper panel of Figure 1. Equivalently, in terms of the *B* and *C* functions in the

lower panel of Figure 1, the condition may be stated as *MB = MC*, or *MW = 0*, i.e. marginal benefits (the slope of *TB*) must equal marginal costs (the slope of *TC*) in the optimum. The above condition gives *n** as the social welfare maximising optimal population of the city. In the lower panel of Figure 1, the product of *n** and *W(n*)* equals the value *TW(n*)* in the upper panel. Welfare is maximal because, at *n**, agglomeration economies are optimally utilised in the production and consumption of private and public goods with respect to agglomeration diseconomies.

The issue of city size, however, is more complex. A most important question is whether the socially optimal solution depicted above is actually reached in a system based on individual utility maximisation. Individual decision-making with respect to the formation of a city can be divided into the purely individual choice of residential location, and the collective choice for city size. In point of fact, purely individual migration and collective optimisation may, in certain circumstances, be efficient in a general equilibrium model with a multiple of cities (Atkinson & Stiglitz, 1980, pp. 533-535), but the solutions are not necessarily efficient in a partial equilibrium model of one particular city.

In order to examine the outcomes of individual decision-making and their possible deviations from the socially optimal solution, take the perspective of the individual's maximisation problem expressed in (2) and concentrate on the lower panel of Figure 1. The lower panel is redrawn and somewhat modified in Figure 2.

(Figure 2 here)

Recall that the socially optimal solution occurs at *E*, the intersection of *MB* and *MC* curves. The intuition is that the effect that the last immigrant has on total benefits in the locality must be equal to

the corresponding effect on total costs. Equivalently, the optimum can be found at the point at which the *MW* curve crosses the horizontal axis. This is the point at which the opposite marginal effects, caused by the last immigrant, just cancel out. At the optimal population level $n^*$, the total welfare in the city, described here by the area beneath the *MB* curve and above the *MC* curve, is maximal. By definition, this area equals the area beneath the *MW* curve. An even simpler measure of the total welfare is the area given by the product of $n^*$ and $W(n^*)$.

What would be the market solution given solely by migration? Assume that all individuals are perfectly aware of the benefits and costs in the sites under consideration, and that they are able to calculate and compare all physical and mental benefits and costs commensurably in monetary terms. Assuming also that migration is costfree, possible welfare gains induce spatial evolution. People choose whether or not to immigrate into a city of a certain size. The city under consideration will attract newcomers as long as it can provide better welfare than their original locations of residence. Supposing that the initial population in the city is large enough so that welfare is positive, and that there are other localities in the economy with zero welfare, the locality will attract newcomers up to *P*, the intersection of the *B* and *C* curves. The market solution then is a city of $n^p$ residents. The individual welfare, not only for the last immigrant, but also for all previous residents becomes zero. If the welfare in the other places is higher than zero, migration will continue until the welfare differentials are equalised, and the population solution may be anywhere left of $n^p$. In any case, with continuing immigration, all citizens come to enjoy the benefits and costs of increased population, and the external effects of immigration are internalised into local welfare as is reflected by the curvatures of *B*, *C* and *W* in Figure 2.

The market solution described above is a stable market equilibrium generated by individual gain seeking. Yet this equilibrium is not an efficient one. This is because expansion to $n^p$ evokes total

costs up to point *e* on the *MC* curve, and total benefits up to point *e'* on the *MB* curve. The welfare loss is described by the area *Eee'*, which is equal to *n** times *W(n*)*. The market solution based on free migration is not socially optimal because individual choices are discrete in nature – individuals cannot choose *n* as a solution to expression (2). Individual decisions tend to make the location too big. In order to attain better personal welfare, true optimisation on *n* should be warranted. The optimum solution at *E* in Figure 2 is clearly a planning solution - individuals with free entry to the locality, and being worse off at their initial locations, would rather continue immigration towards point *P*. Public intervention would be necessary in order to ensure efficiency.

Next, consider collective choice, and assume that cities are sufficiently autonomous to decide on their populations. Taking the residents of the city as a collective - or a club - and assuming that they are equipped with preference revealing mechanisms such that facilitate the maximization of the collective's welfare, then *n* is a continuous choice variable in the model. Individuals alone cannot choose optimal *n*'s, but together, as a collective, they can. There are two possible approaches to the optimal choice of the size of the city. Cornes & Sandler (1986, pp. 164-167) call these the within-club approach and the total economy approach.

According to the within-club approach, the optimal size of the city is set to maximise the welfare of the individual residents of the city. In the simplified setting of Figure 2, this is equal to the maximisation of the average welfare *W* in the city. Therefore, optimal population must be set at *n'*, which is the population that maximises *W*. At the culmination point *c*, the optimum condition *dW/dn = 0* satisfies the standard condition of a club theoretic population optimum, namely that the effect of the last newcomer on the benefits of each previous citizen must equal the corresponding effect on the costs side. To put it more formally, *dB/dn = dC/dn*. Implementation of the policy rule means that newcomers are allowed to immigrate until the welfare of every individual citizen reaches its

maximum, at which point entrance to the city is then halted. The cities can exert this kind of power by various direct and indirect measures. These include city planning and housing policies, limits in service provision etc. (Brueckner, 1982.)

The solution of the within-club approach, however, is not socially optimal in the partial equilibrium model of one city. As in Figure 2, it is obvious that the optimum given by the within-club policy rule does not correspond to the social optimum. More precisely, a within-club city remains sub-optimal in population, $n' < n^*$. Because total welfare is not maximised by the within-club rule, a welfare loss measured by the area $n'cn^*$ results. An equivalent measure for the welfare loss can be derived by calculating the geometric difference between the area given by the product of $n^{*\prime}$ and $W(n^*)$ and that given by the product of $n'$ and $W(n')$.

The total economy approach states that the local government should optimise the population of the city so as to maximise the welfare sum in the city. With this approach, it is not the average welfare $W$ but the sum of individual welfares, namely $nW = TW$ in the present model, that is to be maximised. Therefore, the total economy approach reproduces the above efficient solution at $E$, and the corresponding choice of $n^*$. Note that since entrance to the city is closed only after the $n^*$th immigrant, a decline in individual welfare is caused compared to the above within-club case.

The optimum of the within-club case implies a smaller city than the socially optimal solution of the total economy approach. The result does not depend on the assumption about the relative shapes of the $TB$ and $TC$ functions, and it only reflects the fact that total welfare is the product of population and individual welfare, $TW = nW$, and that the loss in individual welfare is more than compensated by the increasing number of residents up to $n^*$. Thus the collective as a whole benefits from extra members while the individual members suffer.

As far as one city is concerned in a partial equilibrium context, the analysis shows that a pure market solution based on free mobility cannot end up with an efficient outcome. Furthermore, the within-club solution yielded by individual welfare maximisation also remains sub-optimal. Only a city that takes the total economy approach will end up with an efficient solution. Therefore, need for a social planner taking the total economy approach appears to be inevitable.

## 6 Conclusions

The paper presents a club theoretic model of the interdependence of city size and welfare in the city. The interdependence is explained by agglomeration economies and diseconomies attached to city size, which internalise into people's benefits and costs of city life. People respond with migration to possible differences in welfare provided by different cities.

The analysis shows that free migration of people alone cannot constitute a market mechanism that would secure optimal formation of a city. The city itself must act as a market agent and determine its optimal population. In operating this function, there are two competing criteria of optimality, the within-club approach, aiming to maximise individual welfare of the residents of the city, and the total economy approach, aiming to maximise the welfare of the city as a whole. These criteria yield different solutions: in terms of efficiency the total economy approach is superior to the within-club approach.

The main results have already been recognised in the literature, but they are interesting from the viewpoint of the role of local governments. The results emphasise the fact that optimisation of population is a vital function of local governance. Such a function has traditionally been banned at

least in Finland. Furthermore, the results suggest that the city should take the total economy approach and maximise the total welfare experienced in the city even though it lowers the individually perceived welfare levels. It might, however, be argued that the within-club approach is more relevant in practice than the total economy approach. This is due to four reasons.

The first reason in favour of the within-club approach concerns the ability to observe welfare. Operating the total economy viewpoint necessitates calculations of total welfare and/or marginal benefits and costs. These concepts are far more abstract than the average welfare and average benefits and costs, which are observed individually by the citizen themselves and revealed by polls and by the systematic pattern of migration. Since the operation of the planning-oriented total economy approach suffers from severe data problems whereas the within-club approach can be based simply on everyday experience, it is quite obvious that the latter is chosen in practice.

The second reason in favour of the within-club approach is theoretical, namely consistence and coherence of decision-making. The market mechanism behind the formation of a city is based on individual utility maximisation. Since migration between cities derives from purely selfish reasons, it would be absurd to request that, when entering to the city, people should change their perspective from a selfish orientation towards an altruistic total economy approach. This would imply that while they make their private migration decisions according to one optimisation problem, they should, at the same time make collective decisions by maximising another utility function. Since the solutions to these functions are different, such a change in behaviour would be inconsistent. Furthermore, the total economy viewpoint is incoherent with as the key feature of club theory, namely that the formation of clubs is driven purely by the market mechanism based on the individually reasoned choices that people make both in private and collective contexts.

Third, it can be argued that the within-club approach is implicit in the system of representative democracy, which is the usual base in practice for local governance. Given that people are consistent in their utility maximisation, they will vote for those decisions/representatives that promote their personal welfare. Candidates campaigning for the total economy approach would simply not be elected.

The final argument in favour of the within-club approach is voting with the feet. If the local government were to ignore the preferences of the voters by taking a total economy approach, the voters would still have the final word. If there are other locations that offer better welfare, the induced emigration will eventually nullify the social planner's optimisation. The total economy type solution thus is not necessarily stable.

To sum up, the question of the optimal size of a city seems problematic in theory. However, the evidence of the above model of one particular city is less rigorous in practice than in theory. On the one hand, it must be noted that the alternative optimums discussed above are either at the top or along the falling regime of the individual welfare curve. Thus the city is big enough for the net effect of agglomeration economies and diseconomies to be either zero or negative. In practice, however, most cities operate on the rising parts of the welfare curves, and they are in no position to consider optimisation of population by the above welfare grounds. On the other hand, while some solutions may be inefficient in the partial equilibrium model of one city, both the purely migration-driven market solution and the within-club type collective solution can, in certain conditions, be socially efficient in a general equilibrium with a multiple of cities.

**References:**


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

Brueckner, J. (1982): A Test for Allocative Efficiency in the Local Public Sector. *Journal of Public Economics,* 19, p. 311-331.

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. (1989): *Urban Economic Theory. Land Use and City Size.* Cambridge University Press, Cambridge.

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

Krugman, P. (1991): *Geography and Trade.* MIT Press, Cambridge Ma.

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.

Ng, Y-K. (2000): *A Case for Happiness, Cardinalism, and Interpersonal Comparabiliry.* In: H. Dixon (ed.), Controversies in Macroeconomics, Growth, Trade and Policy. Blackwell Publishers Ltd., Oxford.

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)

Scitovsky, T. (1954): Two Concepts of External Economies. *Journal of Political Economy* 62, April.

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

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

Figure 1: Benefits, costs and welfare generated in a city

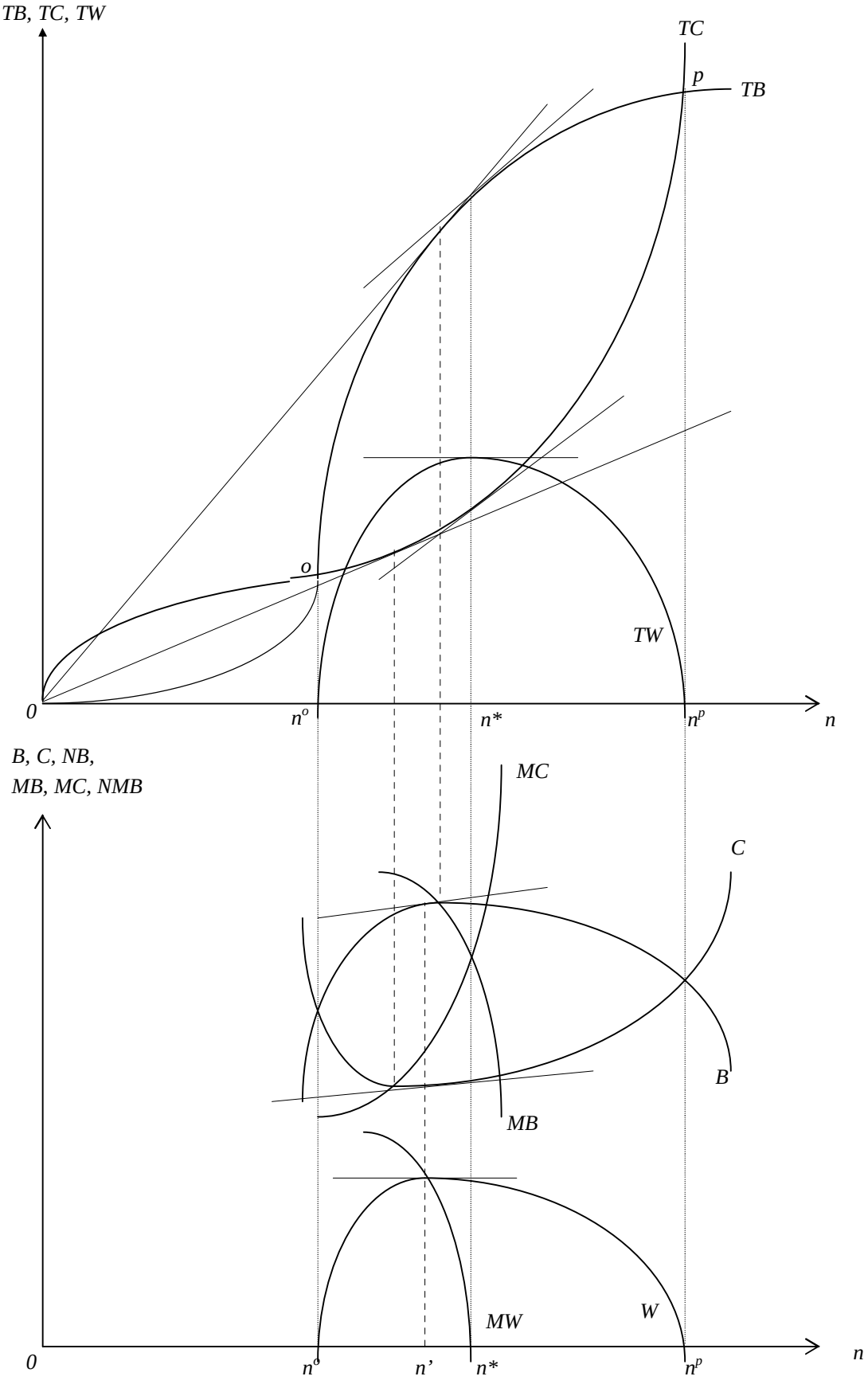
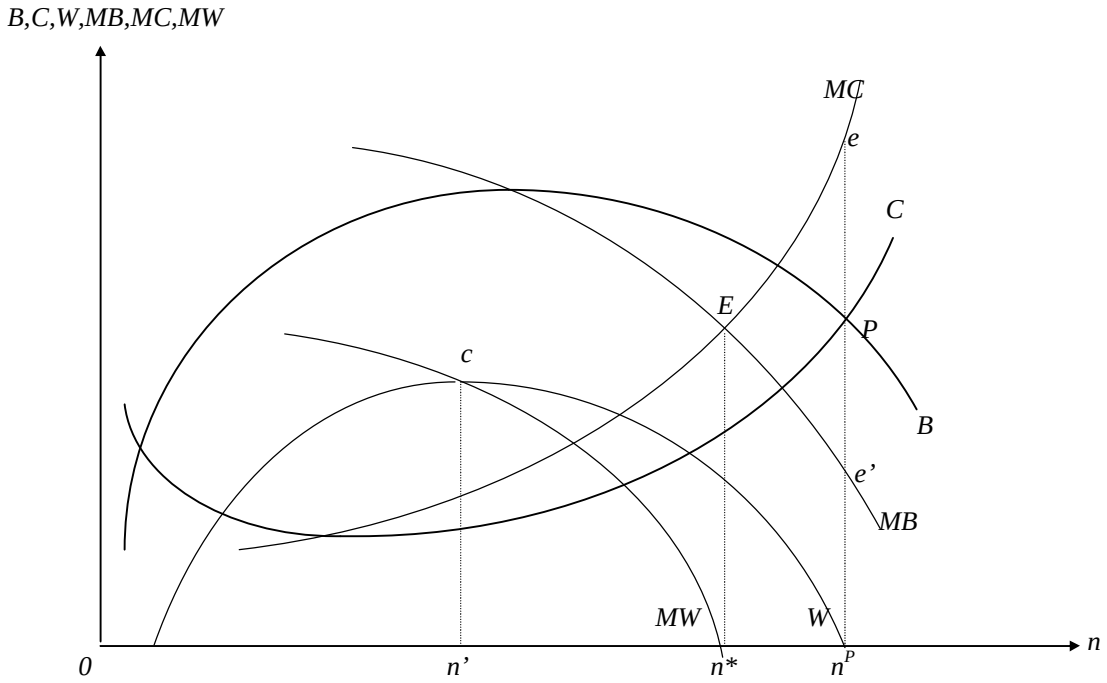


Figure 2: Three possible solutions for city size



Essay #5: Migration, Policy, and the Formation of Cities

***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 club theoretic analysis of migration between cities. It is shown that policy interventions are necessary to ensure efficiency. The results of the effects of policy somewhat contradict the conventional wisdom. First, administrative and economic policy measures are shown to differ in their effects, and second, if cities are initially outside their within-club optima, lump sum transfers accelerate rather than stabilise spatial evolution.*

*Key words: agglomeration economies, city size, exit, migration, voice*
*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 it is understood that economic resources are 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 govern the formation of the spatial economy. Location in general and urban location in particular is a key determinant of profits and welfare. A well known fact is that economic development and urbanization go hand in hand. On the one hand, this is because everyday transactions in the factor and goods markets take space and necessitate transit. Proximity of people and economic functions lowers transaction costs. On the other hand, there are economies of scale and scope, economies of localization and urbanization, and other externalities that have direct and indirect effects on profits and welfare. These agglomeration economies have become more and more important, and urban surroundings quite indisputably offer the most fertile soil for modern market economy.

This paper examines the working of the market mechanism in producing an efficient system of agglomerations, hereafter called cities. A club theoretic framework is used to provide a general equilibrium treatment of spatial evolution of the economy. For the sake of simplicity, the location decisions of firms are taken to be subordinate to people's choices. The firms are assumed to react instantly to market changes so that they are always in their profit maximizing locations. Focusing on migration of people enables a simple translation of efficiency into terms of social welfare, reflecting the spirit of the market economy that the production factors, owned by the people, must be used so as to maximise people's utility from consumption of the final goods.

The club theoretic framework focuses on three issues that are of particular interest here, namely those of the size, formation and governance of the city. The size of the city is important, because it reflects the above mentioned connection between proximity and welfare. Welfare in a city is a product of private and collective consumption and production activities, taking place on local goods and factor markets. The concept of agglomeration economies together with the inseparability of local public goods and services connect city size to welfare experienced in the city: an increase in the size of a city yields benefits and cost savings, but eventually it may also lead to falling benefits and increasing costs.

The formation of cities is basically driven by the migration of people, who seek for optimal locations 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 have on other individual agents. Free mobility of people based on perfect information about the benefits and cost attached to different cities creates the purely private element of the spatial market mechanism.

The issue of governance arises from the fact that peoples' location decisions are not continuous in nature. This so-called non-convexity problem is that people cannot choose the optimal amount (city size), as they do in conventional markets. However, since city size is a determinant of welfare, it must also be an endogenous choice variable. Governance within the city solves this problem: the residents of the city exercise their choice of city size as a collective, through local democracy.

Free mobility and collective optimization of city size should together be able to provide an efficient market mechanism, at least if there are no inter-city spillovers. This paper examines this club-theoretic intuition in detail. The paper proceeds as follows. Section 2 starts with a textbook case of migration between big and homogenous cities and illustrates the welfare effects of collective optimization on the population. Section 3 examines the reverse case of small cities. Section 4 introduces the complications caused by various kinds of asymmetries, and investigates the need, instruments and effects of centralized policy. Section 5 concludes.

## 2 Big homogenous cities

### *2.1 Exit*

The standard textbook case concerns an economy that consists of a multiple of large 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.

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, who make the most preferred choices among the alternatives. Call this kind of a choice the *exit* decision (Bailey, 1999). For the sake of simplicity, fix the total population in the economy at $\underline{n}$ and the number of cities to $\underline{m}$. Thus 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 and diseconomies (Richardson, 1973). The welfare that a city can offer its residents can thus be presented by an inverted U-shaped average welfare curve. It must be noted that, since the relevant measure for welfare is that experienced by individual migrants, the full shape of the curve is unobservable – the only relevant observation is the value of average welfare at a certain point on the curve. The average welfare curve is accompanied by a marginal welfare curve, which remains abstract to the people. Figure 1 presents these kinds of schedules superimposed on each other in a pair-wise setting for two cities.

(Figure 1 here)

In Figure 1, there are two sites of residence, a 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 corresponding 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 $N$. (In a simpler exposition with only two cities $N = \underline{n}$.) 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 to be homogeneous. A less strict interpretation is that the schedules reflect (ex post) systematic responses of people. Some people may migrate against the mainstream, but this can be 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, 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 to the right 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 as beneficial, that is, until the average welfare in both cities becomes equalised. The market equilibrium occurs at the intersection point *e*, where

(1) $W_A^e = W_B^e$.

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 is equal in every city, that is

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

The 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 the perfect homogeneity of the cities, the marginal welfare curves $MW_A$ and $MW_B$ intersect at point $E$, which lies 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 note 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 also include the welfare 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

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

The conclusion is that the market solution is efficient because social welfare is maximised. Therefore, individual utility maximisation ends up at a social optimum that gives the maximum social welfare attainable for all the $\underline{n}$ people in the $\underline{m}$ localities.

### *2.2 Voice*

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). Thus the collective power within 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 thus overcome. Assume that the collective preference revealing mechanism within the cities is efficient.

In the optimization of their population, the cities can take a total economy approach and aim at maximisation of the total welfare in the city, or they can take the within-club approach and aim at 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 $n^1$ is actually such that maximizes total welfare in city B. This is because the $MW_B$ curve strikes through the horizontal axis at $n^1$. However, as 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 the doors against additional immigrants, a city can protect itself from the negative effects of migration. Implementation of such a policy, however, may be difficult unless the initial starting point of migration is on the rising part of the $W$ curve in the city considered. If so, the policy is operational. The situation depicted in Figure 1 is more academic in nature, because the implementation of the policy in that case would necessitate the expulsion 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 deteriorate. This problem can be overcome by letting the number of localities *m* be variable. In the longer term, new cities may enter the city system, i.e. the number of cities is endogenous.

Assume that *m* is variable and let the cities optimize on their population 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 pairwise setting of Figure 2. At the equilibrium point $e^*$ the marginal welfare schedules also intersect. Ignore the integer problem and assume that all people are included in the resulting $m^*$ optimal cities so that

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

The result is that the welfare of all residents in all cities is maximized and equal. 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, all of which maximise the welfare of their residents. Social welfare is thus maximised, too. 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 other *m*-2* cities in both versions). Recalling expression (2'),

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

Any other allocation than $n^{e*}$ would yield lower social welfare. 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 $W_i^* > W_i^e$, thus all people are now better off since they are getting the maximum attainable welfare in the more numerous cities in which the superior welfare is experienced. Being at the top of the average welfare curves in each city means that agglomeration economies are optimally utilised with respect 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.

### *2.3 Policy considerations*

The standard case of large homogenous cities yields a reassuring result: efficiency is guaranteed with or without collective decision-making in the cities. Free migration alone (exit) ends up at 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.

The last modification to the analysis is to consider the total population *n* also as a variable so 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. In 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 $N = 2n/m$ states 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 also obvious that *n* may either increase or decrease. On the one hand, the exclusion of people from the optimized cities can result in an outflow from the economy and cause *n* to decrease while *m* increases. The path to equilibrium *e** becomes even more straightforward. The long-term equilibrium is in principle achievable even among the initial set of $\underline{m}$ cities with a smaller number of total population. The negative side of this result is that social welfare falls compared to (3') and possibly also to (3) because there are fewer 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 causes *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 that of the original inhabitants.

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 thus give reason for a centralized policy intervention to stop emigration from the economy.

# 3 Small cities

## *3.1 Exit*

The setting in the section above is constructed on the assumption that the cities are initially large so that the market solutions appear on the falling regimes of the average and marginal welfare schedules. Migration flows from larger towards smaller cities cause agglomeration diseconomies in the whole system to diminish. However, such 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. This case is illustrated in 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 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 $W_A' - W_B' > 0$ opens up. The welfare gap now starts to attract systematic migration to city A. This, in turn, makes the welfare gap even wider, and further accelerates migration. Point *e* is clearly not a stable solution. In point of fact, free migration would continue until city B becomes totally deserted. If the stochastic element of migration had emerged towards B, then A would have been deserted. The two stable market equilibria on the left or right vertical axis are corner solutions in nature.

The evaluation of efficiency is slightly more complicated in the case of small cities than in the case of large 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 of 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*. Whether or not the corner solutions are efficient compared to point *e*, however, depends on the curvatures of the *W* and *MW* schedules near the vertical axis. In particular, if the *W* curves meet 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 is clearly a need to prevent deviations from point *e* type non-stable but at the marginal efficient solutions. (Atkinson & Stiglitz, 1982, pp. 533-535.)

### *3.2 Voice*

The analysis of Figure 3 reveals that there is a need for intervention. The next step then is to ask if the intervention could be local in nature. Investigate the effects of local policy assuming again that the cities apply the within-club type optimization to their population and letting *m* and *n* be variable. For city A in Figure 3 local policy-making implies that the welfare increasing immigration is stopped at the highest point of $W_A$, i.e. $n^A$ is chosen. In this case the policy is easy to implement – it suffices to close the 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 the doors, but some less fortunate cities continue to lose 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 axes, 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 at $e^*$, 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

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

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 also superior to (3") for the whole set of initial cities $\underline{m}$.

### *3.3 Policy consideration*

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. In general it cannot be ruled out that the total population may 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 compared to (3").

# 4 Asymmetric settings

## *4.1 Exit*

In the above two applications of the model it is assumed that the cities are not only large 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 reveal the essential elements behind the evolution of the city system. In practice, however, the assumption is seldom satisfied. Both components of the assumption need critical assessment.

In practice, any system of cities usually consists of a few large 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 states 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). Obviously, notable migration flows occur between cities of different sizes. Migration flows can also occur between cities in different positions as to the utilisation of agglomeration economies. In the rank order of the cities, the largest cities (and often the most attractive ones) may suffer from dominating agglomeration diseconomies, while the smaller ones at the same time enjoy dominating agglomeration economies.

Second, there is good reason to believe that the cities are heterogeneous in their capability to create welfare due to variation in geography and climate, natural resources, national infrastructure and networks, transport and trade connections, industrial structure, administrative status etc. Heterogeneity results in differences in the respective sets of the *W* and *MW* schedules. The role of the assumption is evident on 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 lie vertically one

upon another. Thus the stable solution of free migration is not necessarily efficient even in the short-term case of large 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 the sake of simplicity, consider two types of cities, type A and type B, of which the A type cities have an absolute advantage in generating welfare, namely 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 both *m* and *n* are 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 at two points, *a* and *b*, which suffices to illustrate all kinds of pair-wise situations between cities A and B.

Start by analysing 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. At the intersection point *a* welfare in both cities is equal, but it is not stable market solution of free migration. To the left of point *a*, a welfare gap opens in favour of city B, and migration from A to B draws the solution further to the left. Right from point *a*, a welfare gap opens in favour of city A, and migration draws further to the right.

Around the intersection point *b*, at the other end, the more modern but excessively congested city A encounters competition for migrants with a traditionally equipped small city B. Now the welfare

equalizing solution at point *b* is stable. To the left of point *b* city A attracts immigrants from city B drawing the solution to the right. The opposite is true to the right of point *b*. Thus, point *b* is a stable market equilibrium generated by migration

Neither of the welfare equalising points *a* and *b* in Figure 5, however, is socially optimal. Not even the stable solution at point *b* is efficient – any other allocation of population to the left of $n^b$ would clearly imply greater social welfare. The superior welfare potentials of the cities remain unexploited because people are only capable of perceiving the existing welfare differentials. The socially optimal solution occurs at point *E* at $n^E$, where the relevant *MW* curves intersect reflecting optimal utilisation of agglomeration economies in the city system. Therefore, for example at the stable equilibrium point *b*, city A is overcrowded and city B is under-populated compared to optimal utilisation of agglomeration economies in the economy. Migration alone clearly cannot lead to a socially optimal solution.

### *4.2 Voice*

Next, consider what might be the result if local optimisation of the population is assumed. Recall that Figure 5 is drawn on the assumption that both cities A and B optimise their populations according to the within-club rule and *m* and *n* settle to their optimal values and no integer problem exists. The outcome is then that depicted by $n^*$ in the figure. 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 $W_A^E - W_B^E > 0$.

In terms of welfare, the result of local level policy-making reads for the pair of cities

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

Through expression (4) it is obvious that achieving the social optimum is not possible without a centralized policy intervention.

### *4.3 Policy considerations*

The two standard instruments for a centralized policy are administrative instruments and economic instruments. The analytical difference between administrative and economic instruments is that administrative instruments operate along the horizontal axis, and economic instruments operate along the vertical axis in Figure 5. The conventional wisdom concerning the instruments is that their effects should be horizontally and vertically equal. Therefore it should be expected that both instruments should treat the market solution equally.

The administrative quantity rationing uses legislative and other such policy measures to make the optimal allocation $n^E$ sustainable. Implementation of the policy means that people are allowed to move from city B to city A until the allocation $n^E$ is reached. In this planning solution the welfare difference $W_A^E - W_B^E > 0$ between the cities is restored. The policy is feasible if the instruments for the optimization of population are in the hands of the central, not local level. Nevertheless, social welfare in the whole economy is maximal,

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

In the above policy version, the economy is divided by policy into better-off and worse-off cities, thereby ignoring the principle of regional equity. However, if regional equity is to be respected, the economic instrument should be used. The central government can use cash transfers to level out the welfare differences and thus make the 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. Figure 6 illustrates the effects.

(Figure 6 here)

In Figure 6, lump-sum transfers from city A to city B shift of the $W^A$ and $MW^A$ curves vertically downwards and the $W^B$ and $MW^B$ curves vertically upwards, leaving the shapes of the curves unchanged. As a result the optimality condition given by the intersection point of the *MW* curves shifts to the left from $n^E$. Transfers draw the two sets of curves to intersect at point $E^*$ at the allocation $n^*$, which is the same long-term allocation produced by local policy-making. The solution is both welfare equalising and efficient. More formally

(4") $SW^{E^*} = n^* W_A^{E^*} + (N-n^*)W_B^{E^*} = SW^E$.

Recalling expression (4), it can be concluded that centralized policy intervention either by administrative or by economic measures not only ensures Pareto efficiency, 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 thus equal. However, in terms of inter-city welfare and, consequently, in terms of allocation of population, the effects are different.

The above findings are interesting for two reasons. First, they contradict the conventional wisdom that administrative and economic measures have equal effects. The effects are shown to be equal only in terms of total welfare but not in terms of inter-city welfare and allocation of population. Second, the findings contradict the common understanding that inter-regional transfers should, as they level out the welfare differentials, stabilise migration. A lump-sum transfer system should be neutral as to individual choices. The main result in this respect is that lump-sum transfers from

richer to poorer cities actually imply that the efficient residential pattern is not the same as that without transfers. That is, if cities are initially outside their within-club optimums, lump-sum transfers rather accelerate than stabilize migration. Particularly, a stable market optimum such as point *b* in Figure 5 is by the policy caused to shift leftwards towards point E* in Figure 6.

The above results concerning lump-sum transfers are derived on the assumption that the policy aims at welfare equalisation by transferring from city A to city B. In practice, inter-municipal grant systems often aim only to level out the differences somewhat, and the grants may not literally go from better-off to worse off places. Taking this into account does not, however, change the qualitative results – the only difference is that condition (4") does not necessarily hold in this case.

### *4.4 The Tiebout hypothesis*

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 doctored by centralized transfer policy. The result is due to the assumption of heterogeneous cities but more or less homogeneous people. Elaboration of these assumptions lends support to the working of the spatial market mechanism, at least in principle.

As to the assumption of heterogeneous cities, it can be argued that in the longer run the initial circumstances in the cities would change and the cities would converge and become increasingly homogenous. The possible solutions would then also converge with 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 that convergence is at best likely to remain partial.

Relaxing the assumption about the homogeneity of people and letting them be different in their experiences of welfare yields a more powerful argument for the existence of a market mechanism. The famous Tiebout (1956) hypothesis states that if people are heterogeneous and free to migrate, they can form an effective set of cities, internally homogenous but externally heterogeneous. As the size of each such city is optimized by maximization of average welfare, social welfare in the economy is maximal. If there are no inter-city externalities the solution is based solely on the *exit* and *voice* type decisions of people, and there is no role for centralized policy.

The Tiebout hypothesis yields an analogy between the club theoretic framework and the competitive market model, which is vital in translating the message of the abstract competitive models into the language of the real spatial world. Due to the importance of the issue, the variants of the Tiebout model are numerous. The existence and efficiency of the Tiebout equilibrium, however, rest on a fairly 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 of an efficient equilibrium distribution of households among cities has been actively discussed in the literature. The treatments in the literature are usually more partial than 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).

Nevertheless, the Tiebout hypothesis is intuitively appealing. Being aware of the obvious existence problems and the strictness of the assumptions, the argumentation can in principle be regarded as

reasonable enough to describe the long-term market mechanism of efficient spatial allocation of people. Unfortunately, the Tiebout case is hard or even impossible to treat graphically because of the added dimension of the heterogeneity of people. The two-dimensional and pair-wise presentation does not suffice to depict city-dependent variability 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 are equal in each heterogeneous city so that nobody has a motive to emigrate.

## 5 Conclusions

The club theoretic model of the paper connects welfare in a city to its size and assumes that there is an optimal size for a city explained by the utilization of agglomeration economies and diseconomies. Inter-city migration based on welfare differentials provides the basic market mechanism for the formation of cities, but it obviously cannot alone allocate people efficiently in geographic space. Choices of location differ from those on common commodity and factor markets in that people cannot choose the preferred amounts or visitation rates. 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 the migration of people can yield efficiency only in the special case of homogenous people and large 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 the personal welfare of their residents. The long-run equilibrium necessitates that the number of cities increases so that optimal utilisation of agglomeration economies is reached in all cities. In the case of small cities, migration will lead to corner solutions without local optimisation of population.

A Pareto efficient optimum necessitates that the growing cities optimize on their population, and that the number of cities falls as 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 will be complicated unless a global perspective is adopted.

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

Two options of centralized policy are studied, namely administrative regulation in the form of quantity rationing, and economic measures in the form of lump sum transfers. Administrative regulation can be used to maintain the efficient allocation of people in spite of the existing welfare differentials between more prosperous and less prosperous cities. Such measures may be difficult to implement in a free market economy because they necessitate constraints on mobility.

Transfers between the cities constitute a more applicable policy measure. Lump sum transfers from richer to poorer cities can be used to equalise the welfare differentials. The social welfare effects of administrative and economic policy measures are the same but, because of the welfare equalising nature of the economic measure, the effects on individual welfares and on the final allocation of people are different. In the transfer option people are allocated so that maximal 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, thus fewer 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 show 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. Lump sum transfers actually necessitate reallocation of people between cities compared to the option of administrative regulation, and also compared to a stable market solution based purely on migration. Therefore, lump-sum transfers rather accelerate than stabilise migration, which is a result that contradicts the conventional wisdom that inter-regional transfers should be neutral in their effects on individual choices.

The final remark regarding the question about the functioning of the spatial market mechanism is given by the Tiebout hypothesis. If people see the connection of city size and welfare differently and gather according to these preferences, then a set of heterogeneous cities with homogeneous populations within them 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 practical usefulness of the abstract hypothesis may be controversial, but it still merits attention. Its key point is that an efficient market equilibrium is feasible in the spatial context without any intervention by central government. Of course, it must also in this case be stipulated that there are no inter-city externalities or other issues that would be of national-level interest.

## 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 large homogenous cities

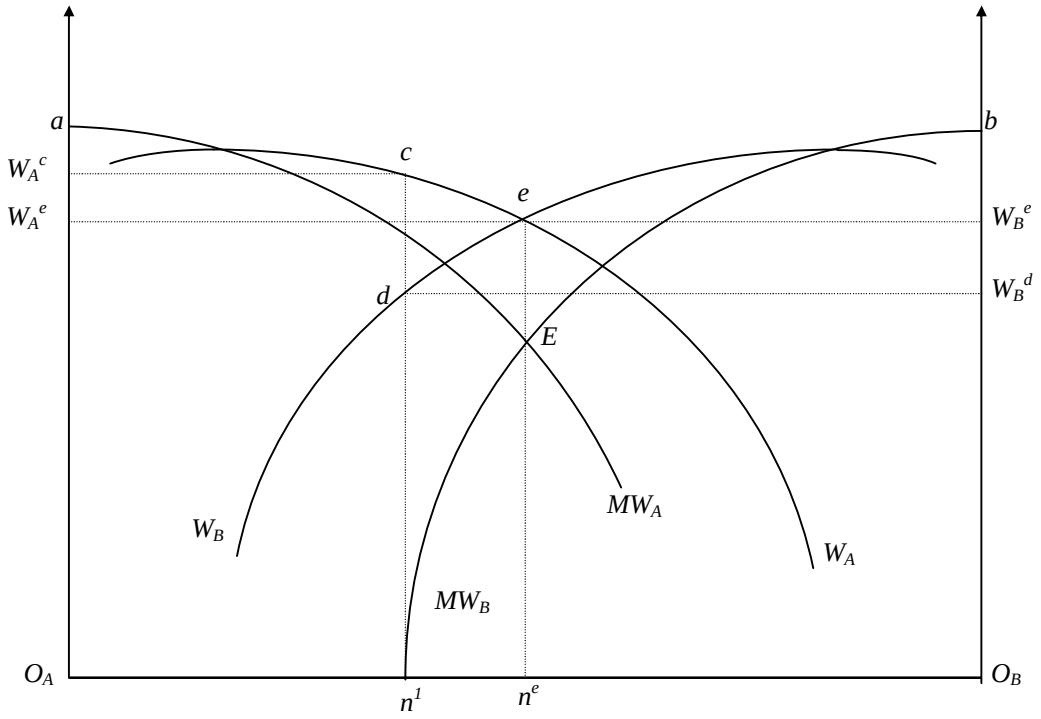


Figure 2: General equilibrium with optimization of city size

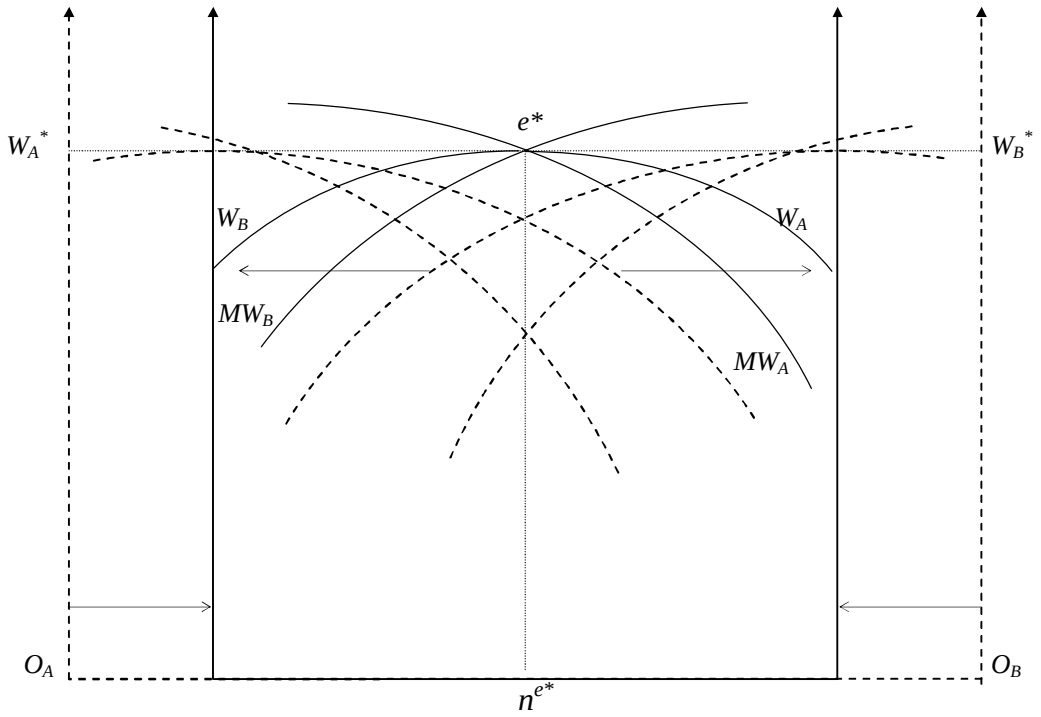


Figure 3: The case of small cities

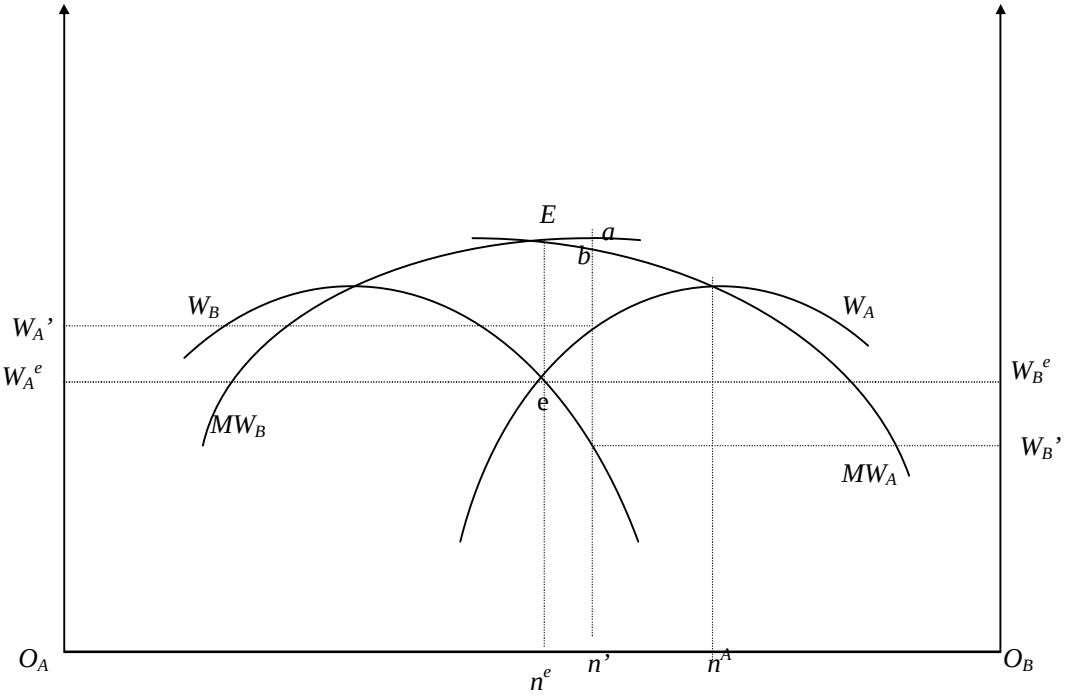


Figure 4: General equilibrium of small cities

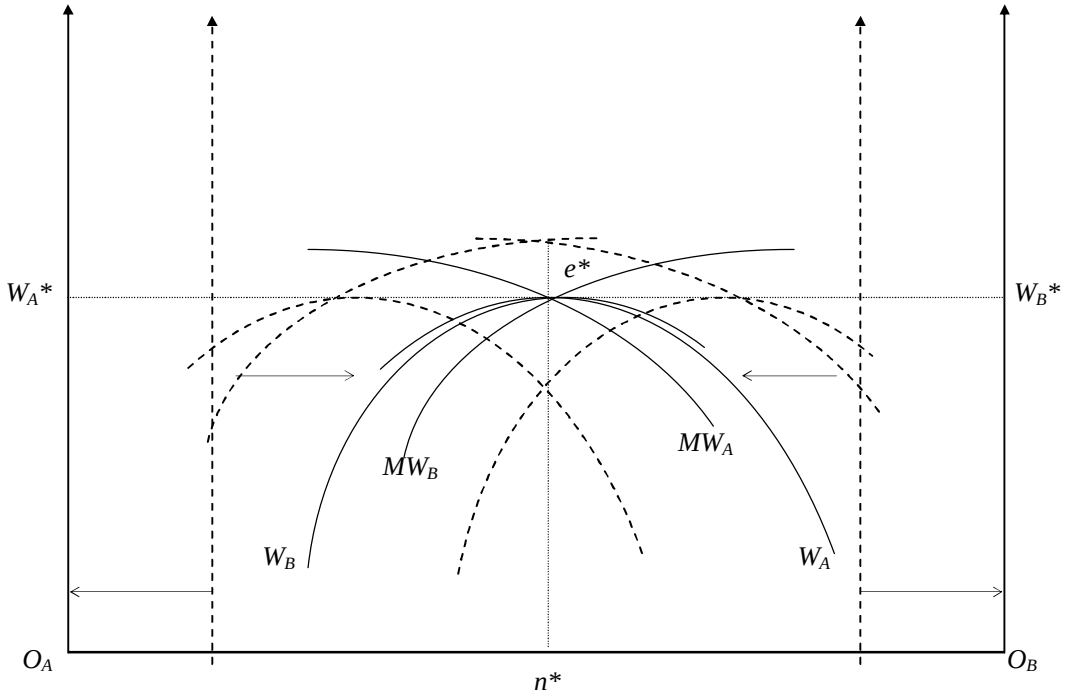


Figure 5: Heterogenous cities

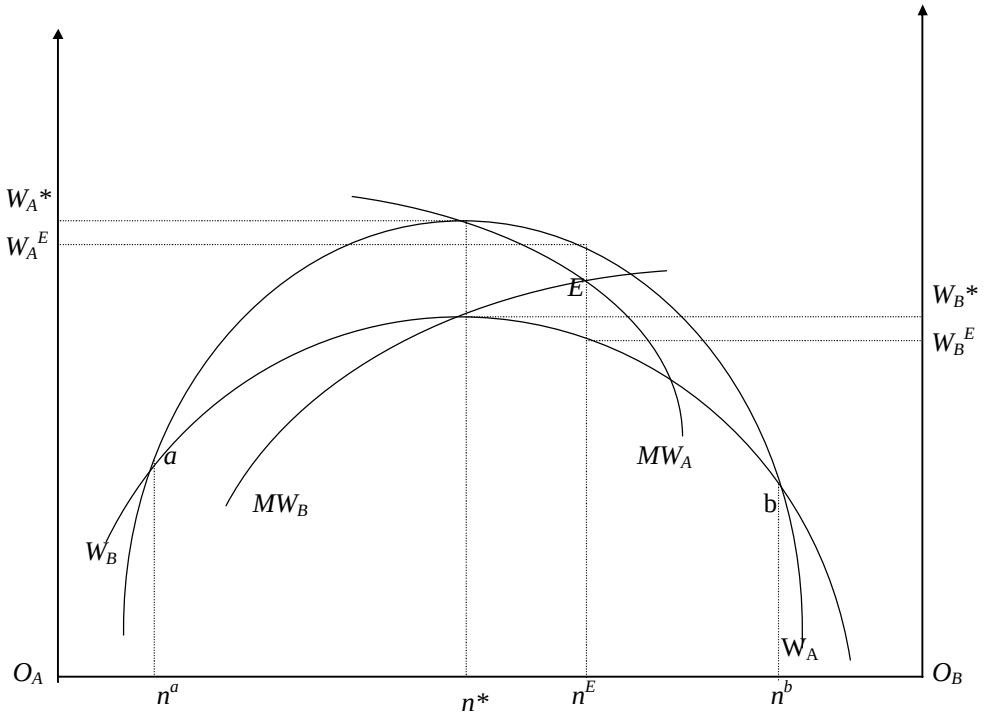
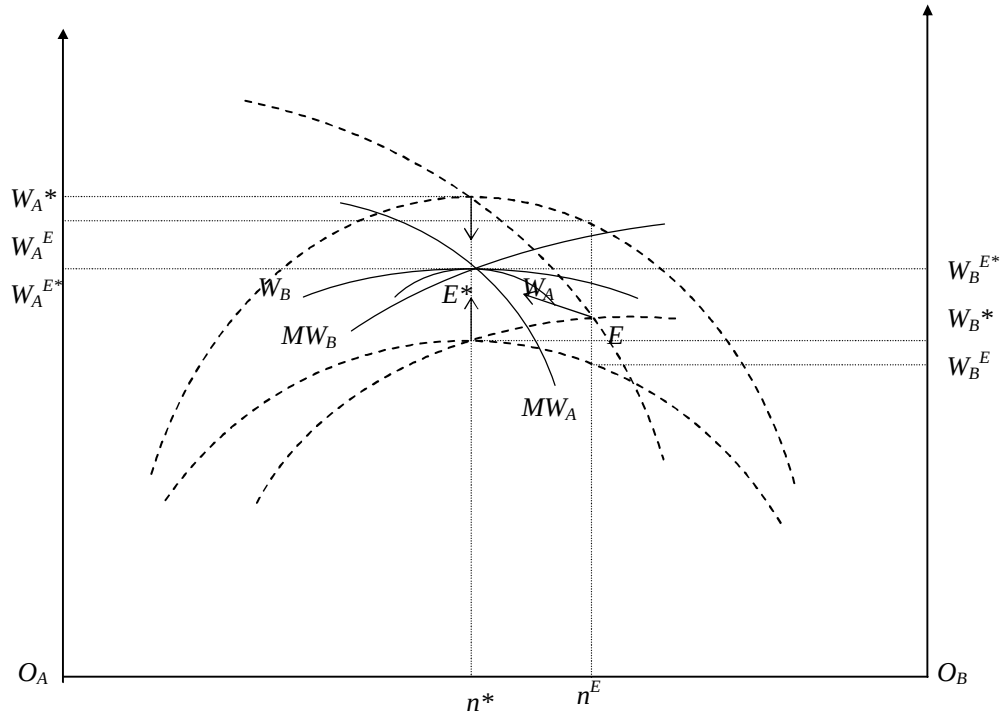


Figure 6: Lump sum transfer from city A to city B



# Essay #6: Urban Governance, Competition, and Welfare

***Abstract:*** *The market mechanism of spatial resource allocation is examined in a system of cities, where social welfare depends on city size. The competitive dynamics of the system is a product of the interplay between people's individual exit type choices (migration) and their collective voice type choices (urban governance). It is shown that the use of efficiency enhancing measures of urban governance depends on the pressure of exit. A necessary condition for dynamic efficiency is that the market equilibrium of migration is non-stable, which sounds somewhat paradoxical. Dynamic efficiency is more likely to emerge between initially small cities, in which agglomeration economies dominate, than between initially large cities, in which agglomeration diseconomies dominate. The incentives for proper urban governance are somewhat ambiguous, however, in the most relevant case, where cities are of asymmetric size. It is therefore important to strengthen the incentives by means of national urban policy.*

*Key words: exit and voice, city size, migration, reactive/proactive policy measures, Tiebout model*
*JEL classification: 931*
*An earlier version of the paper is published in:*
*Urban Studies, Vol. 41, No. 3, March 2004, p. 683-696*

## 1 Introduction

In the past few years internal migration has been lively in Finland as a reflection of globalisation and European economic integration. The still ongoing structural change from a rural economy towards an urban economy has been shifting the locational advantages of the Finnish production sector from rural areas to urban areas. At the present stage of development, the focus of the migration pattern has turned from rural-urban migration to inter-city migration. The fact that market forces now determine the location patterns of firms and people more directly than before, and that the elements of welfare nowadays lie in urban surroundings stresses the need for economic research concerning the concepts of spatial market adjustment, national urban policy and local urban governance.

People and firms are the key market agents of spatial reformation of the economy. This aspect is well covered by the literature on regional economics and urban economics. The market mechanism is based on welfare and profit maximising choices of location, and on agglomeration economies and

diseconomies that connect these choices to community size. However, as collectives of their residents, the communities can also be regarded as market agents in the reform process. The competitive role of the communities is emphasised in the more recent literature on local public economics.

An early version of a comprehensive treatment that combines, at least implicitly, the above-mentioned elements of migration, optimal community size and competitive governance is the Tiebout model (Tiebout, 1956). The model is a well-known representation of competitive public good provision, or to put it more succintly, of efficient allocation of resources in a spatial context. The model is constructed on two cornerstones, inter-community migration of people seeking better local public services, and competition of the communities for the migrants through attractive tax-service packages. The eventual outcome is a Pareto optimal equilibrium, where nobody can gain by moving to another community.

In the original Tiebout model, the role of competition is sketched very briefly to rest on the implicit analogy between communities and private firms as competitive market agents. The focus is on competition for residents between the communities. The Tiebout model is essentially a demand side model that says little about the supply side, and nothing about the production technology of the localities or the political process inside them (Rubinfeld, 1987, p. 174). This paper attempts to elaborate Tiebout's intuition further by exploring the rationale and preconditions of inter-community competition by turning the focus on the competitive pressures that emerge within the communities. The welfare of the residents in a community is taken to be determined by the size of the community on the one hand, and on the actions of the community itself on the other hand. A simple club theoretical framework consisting of a system of conurbations, henceforth called cities, is used. The model is further extended from that of Tiebout by including the outcomes of private

goods and factor markets and the quality of life in people's welfare considerations, and by taking the conditions of the local private markets into the sphere of public influence.

The evolution of the system of cities is studied from the viewpoint of people's choices. Clustering of the firms is an essential source of agglomeration economies, but, for the sake of simplicity, the firm sector is ignored by assuming that the profit maximising choices of the firms respond competitively to the evolution of the local markets, and that they are effectively reflected by the parameters of the budget constraints faced by the people. People make their choices between and within the cities. Between the cities people make their choices by exit, that is, by voting with their feet between residential sites. Within the cities people make their choices by voice, which is to say that they take democratic and other collective actions so as to affect the local preconditions of their welfare. The collective mode of people's decision-making is the origin of what is here called urban governance. It is also assumed that the preferences of the city residents are effectively revealed on the average by means of local democracy, and that they are carried out without distortion by the local bureaucracy.

Exit-type choices are close but not equal to normal market choices. In private goods and factor markets the market agents exert exit between preferred and non-preferred items, and they choose optimal amounts of those preferred. Residential choices are different, because people are not able to carry out the second stage of decision-making – they cannot optimise on the quantity that is the size of the city, which indirectly determines their utility in the city. Furthermore, exit-type choices are made according to average, not marginal concepts, as they are made in the market case. Therefore, exit does not suffice to secure efficient market outcomes in a general sense (Atkinson & Stiglitz, 1982). However, by using their voice the residents of a city are collectively able to optimise on the

size of the city. In collective decision-making, the problem of marginal vs. average choice criteria is also irrelevant (Cornes & Sandler, 1998).

The necessary condition of allocative efficiency is that all market agents operate in a technically efficient manner. As is well established in the standard theory of the firm, competition is the main driver of technical efficiency. Competitive dynamics in the system of cities precludes external pressure to evoke technical efficiency within the cities as in the case of competing firms. At first glance the analogy seems to be straightforward - the threat of emigration (exit) should force the cities to efficient conduct of urban governance (voice) just as firms have to respond to declining market shares with greater efficiency. The situation here, however, is more complicated. In the following the nature of competitive dynamics of the system of cities is examined in close detail. An analogy is sought between the position of private firms and that of the cities.

# 2 The basic model

## *2.1 Agglomeration economies and welfare*

A conventional postulate in the literature of urban economics is that welfare in a city depends on its size (Alonso, 1971). This is because household, business and social agglomeration economies (and diseconomies) arising in densely populated urban surroundings determine both the benefits and the costs of urban life (Richardson, 1979, p. 306). This classification points to the fact that locally enjoyed welfare depends on the efficiency of the activities of the firms and the local public sector in the city. In this paper, the setting is simplified by assuming that the firm sector is competitive and adjusts efficiently to the respective circumstances, and only the private and collective actions of the households are studied.

The optimisation problem of a site-seeking household can be stated as a standard consumer's choice by assuming that the household agglomeration economies are felt directly on the benefit side (in the utility function), and that the business and social agglomeration economies are faced indirectly on the cost side (in the budget constraint) (cf. Richardson, 1973, pp. 11-20). The direct benefits of urban location are anticipated in everyday life, and the arguments of the utility function include quantities of private goods, public goods and leisure. The quality of life can also be included in the leisure variable. Agglomeration economies enter as technical externalities. The costs of urban life, described by the budget constraint, are determined by local factor prices, particularly wages on the income side, and by commodity prices, transportation costs, local taxes and tariffs etc. on the expenditure side. Here agglomeration economies work indirectly through the price mechanism, and are thus pecuniary in nature. Because of agglomeration economies, both the benefits and the costs are implicitly determined by the size of the city. (Mills & Hamilton, 1984; McCann, 2001.)

In the spirit of constrained utility maximisation the welfare anticipated by an average citizen is determined by the net sum of benefits and costs of urban life. A standard assumption is that agglomeration economies dominate the formation of welfare in the earlier stages of city growth, but that agglomeration diseconomies gain strength with growth and eventually gain ascendancy (Richardson, 1973). Figure 1 illustrates the basic assumptions.

(Figure 1 about here)

In Figure 1, city size is measured by the number of people (denoted by $n$) within the geographical area of the city. This is reasonable if the geography is taken as given in the short term. In the figure two sets of curves are presented. The curve $AW^p = W^p/n$, where $W^p$ is potential total welfare, plots

the potential average welfare consisting of the net sum of benefits and costs in the city considered. The term potential refers to full utilisation of the existing welfare creation 'technology' of a city. The technology itself is taken as given and innovations that change it are ignored. In this sense, the $AW^p$ curve resembles the curve of the attainable profits of a competitive firm, or the production possibility frontier of an economy. The curve $MW^p = dW^p/dn$ plots the potential marginal net benefits that new migrants bring to total welfare in the city. The $MW^p$ curve strikes through the $AW^p$ curve from above at $n^o$, where the latter reaches its maximum point. At that point, agglomeration diseconomies start to dominate over agglomeration economies.

The other set of curves in Figure 1, namely *AW* and *MW*, plots the actually observed average welfare in the city. Due to market distortions and excess transaction costs in the local private sector and inefficiencies in the local public sector, the actual observations may lie far below the welfare potential of the city (Brueckner, 1982). This reflects inefficient utilisation of the existing welfare creation technology. The observation of welfare is the only relevant information. In Figure 1, at city size $n^1$, the average citizen observes welfare $AW_1$, while the theoretical potential would be $AW^p_1$. Denoting the gap between potential and actual welfare by *g*, the total welfare worth of *g* times $n^1$ remains inexperienced in the city. This is the dead weight loss of remaining below the potential of the city. Furthermore, $n^p > n^o$ would be the optimal population if the full potential were achieved.

Since both the benefit side and the cost side are determined by the size of the locality, the number of the population should be a true choice variable in the optimisation of the households, but it is obvious that the households cannot choose optimal population such as $n^o$ or $n^p$ in Figure 1. Therefore, the maximisation problem for an average household stated above actually is the problem of the social planner. For this reason (among other things) the city must now also be studied from the perspective of collective choice.

### *2.2 The role of urban governance*

Urban governance concerns a wide range of public activities that affect the preconditions for welfare in the city (Bailey, 1999). The essential feature here is that urban governance is based on collective decision-making - the citizens make their collective choices by voice, that is, via local democracy. A benchmark case for efficient urban governance is the following: the distribution of welfare in the city is symmetrical, the local public expenditures are financed by lump sum taxes, and the local bureaucracy operates without bias. Then, the welfare of the median voter represents average welfare, and a simple majority-voting rule produces a Pareto-optimal outcome. It is obvious that these features are quite specific in nature. (Cullis & Jones, 1998, pp. 78-87.)

Urban governance can be divided into reactive and proactive functions. In the reactive sense urban governance is operated on a short-term basis with the prevailing circumstances taken as given. The focus is on the observed changes in the average welfare - systematic migration into the city may increase or decrease the welfare of the existing residents. From this perspective, the main function of urban governance is the optimization of population. In Figure 1 population $n^o$ is chosen according to the AW schedule as a short-term solution to the consumer's problem depicted above.

There are some limits to the horizontal optimisation of population. Aiming at $n^o$ from left to right is out of compass, because that depends on the exogenous exit decisions of potential newcomers. Aiming at $n^o$ from right to left may also be difficult to implement, at least if the civil and property rights of the current residents are respected. The function can be operated in practice only by allowing entrance to the city up to $n^o$ (or other stipulated population level) and closing entrance at that size. The operative measures include planning, zoning, housing policy and dimensioning of public services. The reactive optimisation of population has been widely discussed in the literature

on urban economics and it is also in common use in practice. The optimisation criteria, however, are often other (say, real estate values) than that used in this paper. (Brueckner, 1982.)

The proactive function of urban governance concerns full utilisation of the welfare potential of the city. In terms of Figure 1 this refers to bridging the gap between potential and actually observed welfare by shifting the observation $AW_1$ vertically closer to the potential $AW^p_1$, and refining the shape of the $AW$ curve to follow that of the $AW^p$ curve. Proactive urban governance takes care of the technical efficiency of the city as an economic unit. Just as in the case of profit maximising private firms in competitive markets, the survival of a city depends on its technical efficiency. The proactive functions of urban governance include competition policy, structural policy and growth policy.

The concept of competition policy refers here to improvements in the efficiency of the local public and private sectors. Recalling the above benchmark for efficient urban governance, efficiency in the public sector must be ensured through decision-making, financing and execution of service production. It is well documented in the literature on collective decision-making that in practice Pareto-efficient outcomes are seldom reached by majority vote. The decision-making methods must be developed to ensure that urban governance is properly steered by the preferences of local residents. The promises of new technology have received attention in this respect. In public finance, distortive proportional and progressive taxes are much more familiar in practice than undistortive lump sum taxes. Refinements of tax systems, applications of marginal cost pricing and charging policies are examples of financial reforms. The executive problems of bureaucracy are well known. Standard means of improving the efficiency of service provision include privatisation, monitoring, deregulation, competitive bidding, purchaser-provider arrangements etc. (Bailey, 1999.)

Proactive urban governance also concerns the efficiency of the local private sector because it has a vital role in the functioning of the local commodity and factor markets (Prud'homme & Lee, 1999). Urban governance can notably affect both the market distortions faced by the competitive firm sector and the overall transaction costs in the city. As a key constructor of local technical and social infrastructure, urban governance can also contribute to the formation of local social capital, which has a catalyst role in accelerating innovation, and a pre-emptive role in establishing a productive economic environment (Kajanoja & Simpura, 2000). On the technical side, the instruments include zoning and city planning and systems of public transport, logistics and networks, which have considerable effects on the time and money costs of market transactions in the city. As to social infrastructure, the instruments include immaterial networks and recreational, cultural and other such facilities.

Structural policy means stepping up and co-ordinating reallocation of local resources from sunset industries to sunrise industries. The main instrument in this respect is strategic programming. Traditional measures include focusing of local budget funds and external funding in order to evoke economic incentives for warranted development. More modern modes of local structural policy aim to amalgamate the mutual interests of the community and local private business life into joint efforts of strategic development (Bailey, 1999).

Growth policy is about setting long-term size (or growth rate) targets. As it is the density of population that determines agglomeration economies and welfare, both the geographical area and population of the city must be controlled. Regarding geographical optimisation, recall that the *AW* schedules in Figure 1 are drawn for a given area. For areas of different sizes the curves may look different. If economies/diseconomies of scale exist, the *AW* curve can be made to reach a higher peak value by widening/narrowing the city area. The area must be chosen so that the long-run

average welfare is maximised. Instruments of geographical optimisation include municipal consolidations, federations, and regional co-operation in service production and development. Long-term optimisation of population is then about choosing welfare-maximising population according to the proper $AW^p$ schedule. The same operational notions apply as in the reactive mode of urban governance, and the instruments are mostly also the same. As demonstrated later, proper fulfilment of this function is vital for the efficiency of the whole system of cities.

To conclude, voice-driven urban governance is operated under the pressure of exit-type migration decisions. Reactive optimisation of population is a response to inward migration. It is easy to safeguard the welfare of the original citizens by stopping planning for new housing and services just before the desired city size is reached. The decision is politically neutral, at least if the preferences of the median voter equal those of the average household. In the median voter model the outcome is also welfare maximizing in the short term. Reactive urban governance resembles firms' short-term optimisation on production according to market parameters. The effort of applying the proactive functions needs pressure from the threat of outward migration. This is because the implementation of these functions affects the current residents of the city. Equity considerations arise because more efficient tax and transfer policies treat people unequally, structural reallocation of resources causes unemployment in some industries, changes in city planning and public transport systems capitalise on real estate values, and so on. The likelihood of political conflicts is high, and the costs of collective decision-making increase (Cullis & Jones, 198, pp. 82-83). There is again an analogy to private firms: the powerful example of the dynamic inefficiency of a monopoly suggests that such of internal renewals are accepted only under considerable market pressure (Bailey, 1999, pp. 98-100). Let us now make the pressure of inter-city migration exit explicit in the analysis.

# 3 Inter-city migration and competitive dynamics

## *3.1 Large cities*

Assume first that the economy consists of a system of initially large and homogeneous cities. Let the number of cities be *m*, and let the total population be *N*. People are totally mobile and fully informed about the welfare observations everywhere, and migration between cities emerges if welfare differentials are found. This is a standard assumption, and the introduction of mobility costs and uncertainty in welfare comparisons would not change the basic intuition. Figure 2 below illustrates this standard setting. In the figure, the *MW* curves as well as the $AW^p$ and $MW^p$ curves are ignored for the sake of simplicity.

(Figure 2 about here)

In Figure 2 two of identical cities, city A and city B, are presented. The curves $AW_i$, $i = A,B$, plot average welfare in the cities, drawn from left to right for city A and from right to left for city B. Since all cities are identical the length of the horizontal axis is the share of the two particular cities A and B of the whole population *N*, namely $n = n_A + n_B = 2N/m$, where $n_A$ and $n_B$ denote respectively the number of residents in A and B. The households make their choices according to observations along the $AW_i$ curves.

Assume that the total population is initially divided so that the number of people in A and B are respectively $O_An^1$ and $O_Bn^1$. At this stage, there exists an observed welfare gap, denoted by *g*, in favour of city A. People are induced to move from city B to city A. Migration narrows the gap due to two effects. First, migration into A compels agglomeration diseconomies in A, which reduces

benefits and raises costs, causing average welfare in A to decline. Second, migration from city B decreases agglomeration diseconomies in B, which improves welfare for those who remain in B. Continued migration leads to a stable equilibrium at point *e*, where welfare in both cities is equalised and no motives for further migration exist. In the equilibrium, the size of city A is $n_A = O_A n^e$ and the size of city B is $n_B = O_B n^e$. It is clear from the figure and from the above definition of *n* that $n_A = n_B = N/m$. The equilibrium is also efficient, because social welfare, consisting of the area $O_A AW_A^e AW_B^e O_B$ is at its maximum. For any other value of n there would be welfare losses. Note that in this individualistic way of welfare measurement the problem of inter-city externalities and other reasons for paternalistic welfare judgements are ignored (Boadway & Bruce, 1984).

However, the market solution of Figure 2 is efficient only in the static sense. Since the $AW_i$ curves in the figure present observed welfare, assumed to remain below the respective potentials, welfare improvement through higher technical efficiency is possible. The question now is whether there is such a competitive pressure in the system to occasion improvements in technical efficiency in the cities as there is in the case of private firms operating on competitive markets. The competitive pressure needed for improvements in efficiency is generated by exit. Since the equilibrium at *e* is a stable position with zero welfare differentials, there exists no such pressure at that point. People are not on the move and urban governance is not forced into inconvenient improvements in technical efficiency. The case of a stable market equilibrium does not seem to imply high efficiency in the dynamic sense.

In order to gain a more profound insight into the dynamic properties of a stable market solution it is useful to examine an alternative equilibrium path recalling the option of static urban governance. Reconsider the starting point $n^1$ in Figure 2, where migration from B to A makes those who remain in B better off but those who originally live in A worse off. Because the initial residents of A

anticipate that in-migration will reduce their welfare, they may be induced to raise their voice at that moment. The simplest thing they can do to avoid welfare loss is to adopt reactive urban governance and deny entry to city A at $n^1$. Now, migration is unable to level up the welfare gap $g$. For city B, the gap can be bridged only by means of proactive urban governance. The residents in city B may raise their voice in favour of developing the use of welfare creating technology. This is illustrated by the upward shift of the $AW_B$ curve to $AW'_B$. It must be noted, however, that city B is not forced to do this – there is no threat of exit because people cannot migrate to city A. Another notable implication is that there is no great urgency for improvements in city A even if the average welfare in B should increase above that of A at $n^1$. This is because migration to B would benefit both A and B.

If city B manages to improve its technical efficiency, the welfare of the whole system will increase somewhat. This is because B moves closer to its potential, and the original residents of B become better off. However, the situation is still sub-optimal if no improvements in technical efficiency in city A have occurred. On the other hand, if city B is not able to improve its efficiency, the welfare of the whole system will remain lower than in the market solution at point *e*. The general conclusion is that the equilibrium at point *e'* is basically stable, just like that at point *e*, and no exit-driven competitive pressure exists to draw the $AW_i$ curves closer to their potential counterparts.

### *3.2 Small cities*

The above analysis confirms the standard result that with free migration there is a case for a stable market equilibrium. However, because of low incentives for improvements in technical efficiency this equilibrium may well remain dynamically sub-optimal. The result holds in the case of large cities, which operate on the declining regimes of their average (and marginal) welfare schedules. It is also worthwhile to analyse the case in which the cities in the system are initially small and

agglomeration economies dominate. This is also the situation that is more relevant in practice, at least in a national perspective. Figure 3 depicts this case.

(Figure 3 about here)

In Figure 3 the setting is changed from that of Figure 2 by assuming that the system now consists of a considerably larger number of cities $m'>m$, while the total population $N$ is taken to be unchanged. Therefore, in any representative city, the number of residents must now be smaller, in other words, the horizontal axis must be shorter with respect to the $AW$ (and $MW$) schedules.

Reconsider the situation of welfare equalisation presented by point $e$ in Figure 3. With small cities, point e is evidently not a stable market equilibrium. Suppose that, for purely stochastic reasons, some households move from city A to city B, and that the allocation of people shifts from $n^e$ to $n^1$. A welfare gap $g^0$ is opened in favour of B. In this case, migration does not bridge the gap - in fact it makes the gap even wider and causes the solution to deviate from $e$ at an accelerating rate. The only stable market equilibrium in this case would be a corner solution, in which either city A or city B, depending on the direction of the initial stochastic shock, will be deserted.

In the non-stable migration pattern reactive urban governance is of no use. Denying entry does not help to stop out-migration from city A, and city B does not want to cut the beneficial in-migration. The incentive structure of urban governance has now changed from that in Figure 2. In Figure 3, migration from city A to city B makes the old and new residents of city B better off because of rising agglomeration economies, but the residents remaining in city A become worse off because of decreasing agglomeration economies. The residents of the out-migration city now suffer, and thus face the pressure of exit. Proactive urban governance is to be considered.

Under the pressure of accelerating exit by their fellow citizens, the more reluctantly moving residents of city A raise their voice in favour of competition policy and structural policy. Figure 3 illustrates how improvements in local efficiency shift average welfare in A up to $AW_A^{'}$ so as to bridge the welfare gap $g^0$. At $e^1$, average welfare in both cities is again equalised. Assume that city A now looks even more attractive compared to city B at the marginal, and migration turns backwards resulting in $e^2$. A welfare gap $g^1$ now opens in favour of city A. People in city B have to respond to this by expressing their desire for efficiency improvements, which results in an upward shift of $AW_B$ to $AW_B^{'}$. If migrants are now attracted to city B, people in city A will again be forced to make further improvements. The process continues in consecutive order. As a result of proactive urban governance in both cities, the welfare levels actually observed converge with their potential counterparts.

In principle, the competition for migrants between A and B will continue as long as welfare gains can be achieved by improvements in technical efficiency. In the long term all cities will go through the evolution illustrated in Figure 3. After reaching their potential welfare curves, the cities can no longer compete for new residents by improving technical efficiency. At this stage, the non-stable nature of the solution appears confusing, and a reasonable question is whether in the long run an efficient market outcome is at all possible.

Proactive growth policy now comes into the picture. All cities optimise on their geographical area and choose their population so as to maximise the average welfare in that area (Ng, 1973). Those cities that manage to attract people allow migrants to enter until average welfare in the city reaches its maximum. After that, entrance is closed. The remaining cities will continue the competition. Some of them will attract people and some will lose them. The number of cities is endogenously

determined in the long run. The mechanism continues until there is an optimal number of cities all having an optimal population. The cities may turn out to be heterogeneous with homogeneous population in them. Ignoring the integer problem, all people in every city enjoy maximum welfare and nobody wants to migrate. The solution is socially Pareto efficient since nobody's welfare could be improved without making somebody else worse off. (Cornes & Sandler, 1986.) The famous rank-size rule (or Zipf's law) of 'natural' size distribution of cities can be regarded to reflect this kind of development (McCann, 2001, p. 79).

The general equilibrium solution above, of course, remains hypothetical and the social costs of the evolution process may be significant in practice. The dismal fact is that some cities are doomed to be deserted and at least some infrastructure must be relocated as far as total population is taken as fixed. The problem of depopulation can be avoided by taking total population as endogenous. The rural sector or other countries ignored in the analysis above can to some extent serve as a population reserve in this respect. But if there is no such effective reserve, social costs may arise. If the indivisible costs of demographic and economic decline fall mainly on the residents of the declining cities, the migration rate may be biased to exceed that of the 'natural' rate of depreciation of the infrastructure. The focus regarding the problem of sunk costs is on the speed of the adjustment.

### *3.3 Asymmetrical cities*

So far the discussion has focused on symmetrical situations on falling or rising average welfare schedules in both cities alike. In practice, there are often cities in both positions at the same time. Another rather restrictive assumption made above is the assumption of the homogeneity of the welfare schedules between cities. For various geographical, historical, institutional, structural and other reasons, the welfare schedules of the cities may well differ from each other even in the longer term. The economic preconditions of the cities are quite unilateral, especially in recently

industrialised countries, like Finland. For practical reasons, it is most reasonable to take these asymmetries into account in the analysis.

Figure 4 below illustrates the essentials of the case of asymmetric cities. In the figure, the *AW* curves for cities A and B are again presented, but city A is now assumed to be initially a more fertile city with higher average welfare around the peak value than in city B. The *AW* curves are drawn to reflect the difference. The *AW* curves intersect at two points, namely at *e* and *e'* in order to demonstrate all the remaining variations of possible outcomes.

(Figure 4 about here)

At the intersection point e in Figure 4, city A is initially large compared to city B. City A is on its falling welfare regime due to high agglomeration diseconomies, whereas agglomeration economies dominate in city B. Just as in the case of large cities above, it is obvious that point e is basically a stable market equilibrium. Consider first the allocations right of *e*. There is a welfare gap attracting people from A to B. The adjustment will result in a stable market solution at *e*. Since migration is beneficial to both A and B, and the adjustment bridges the welfare gap, no notable incentives for the using of voice exist in either city. As to allocations left from *e*, say at $n^1$, there is a welfare gap that draws people from B to A towards a stable solution at point *e*. This would be inconvenient for both A and B. If city A now denies entry at $n^1$, city B can respond by the effort of proactive urban governance. If the effort is sufficient to bridge the welfare gap so that $AW_B^{'}$ is reached, the system ends up at a stable equilibrium at $n^1$. It must be noted that residential reallocations left of *e* would improve social welfare even without dynamic improvements in B. The incentive structure reduces to that of the case of large cities. Incentives for dynamic improvements in efficiency are weak around the stable intersection point *e*.

At the other end in Figure 4, point *e'* presents a more promising case, where city B is large and city A is small. Agglomeration economies dominate in A and diseconomies dominate in B. The solution at point *e'* is non-stable, since starting from *e'* in either direction opens up growing welfare gaps. If, for stochastic reasons, migration should start leftwards from *e'*, both cities would lose welfare. In this direction, the large city B can and most probably will stop the loss of its residents' welfare by denying entry. With entrance to B closed, city A may try to level out the welfare gap by proactive urban governance, but again it is not the pressure of exit that necessitates this. The possibility of denying entry makes *e'* stable leftwards.

If migration should turn rightwards from *e'* in Figure 4, that is, if people should move from B to A, both cities would gain welfare. City A gains because of increased agglomeration economies and B gains because of decreased diseconomies. The case can be interpreted as one in which a small modern city attracts people from a larger city with a less developed industrial structure. Because there is now an ever-expanding welfare gap, the point *e'* is unstable in nature. In city A there is certainly no apparent need for raising voice, but in city B the incentives for urban governance are somewhat more ambiguous. The incentives rise because of the accelerating out-migration and widening welfare gap, but the incentives are impaired by the fact that the welfare of the still remaining citizens is improved in any case.

To obtain an idea of the dynamics, suppose that migration from B to A continues as such rightwards from *e'* until the allocation $n_A$* is reached. At this moment the need for urban governance becomes apparent in city A: city A adopts the reactive mode and denies entry to optimize on its population. This eliminates the pressure of exit, but leaves a yawning welfare gap. Suppose also that even without the immediate pressure of exit city B succeeds in implementing proactive urban governance

and manages to bridge the welfare gap at point *e"* in Figure 4. At this stage, the competitive dynamics of the system depends on the shape of the average welfare schedule of city B. In Figure 4, the upwards-shifted average welfare curve $AW_B^{"}$ has a positive slope at $n_A$*. This means that city B has managed to implement a major reform regarding the effects of agglomeration economies. In this case the solution *e"* is non-stable, and the system moves to a dynamically efficient general equilibrium just as in the case of small cities presented above. This is because, left of point *e"*, both cities would gain welfare from in-migration and are therefore forced to improve their technical efficiency to attract migrants. However, if the $AW_B^{"}$ schedule sloped downwards at $n_A$* (as presented by the broken line version of the average welfare curve), there is no threat of exit, and point *e"* remains a stable and dynamically sub-optimal solution.

To sum up the findings, competitive incentives for improving technical efficiency are generally rather weak in the case of asymmetric cities. The incentives may remain too shallow to force the application of costly and inconvenient measures of proactive urban governance. Dynamic efficiency is likely to prevail only in the case of a genuinely non-stable market solution. Such a situation is possible as a special case, where there is a welfare gap in favour of the initially small city (right of point *e'* in Figure 4). However, dynamic efficiency rests on two rather restrictive preconditions. First, the large out-migration city should improve its technical efficiency even though it gains welfare from the outflow (moving from *e'* to *e"* in Figure 4). Second, the improvement in technical efficiency must be substantial enough to make the effects of agglomeration economies more positive at the margin than those in the rival city (the average welfare curve must be upwards sloping at point *e"* in Figure 4). Thus the optimal size of city B should be far beyond its short-term optimum $n_B$*.

The above observation stresses the importance of national level policy: the incentives for competitive practices must be enforced. Furthermore, it is not only that the larger cities must be encouraged to use proactive urban governance; the focus must be placed on heavy structural reforms that accentuate the effects of agglomeration economies. In the case of an initially large city the necessary reform sounds major but not impossible considering the disposable resources. The large cities must be motivated to grow even larger beyond their short-term optimality considerations. This benefits not only the cities themselves but also the society as a whole, because it facilitates competitive dynamics and, consequently, dynamic efficiency in the city system.

## 4 Conclusions

The analysis complements the message of the Tiebout model of spatial resource allocation by investigating the preconditions of competitive dynamics in a system of cities. In the present model agglomeration economies explain the connection between city size and welfare, and local welfare is maximised by urban governance (voice) under the market pressure created by migration (exit). The dynamic efficiency of the spatial system rests on the interplay of exit and voice. The incentives for dynamic improvements depend on the competitive nature of the system. In this respect, the cases of small and asymmetric cities are insightful.

Urban governance is divided into reactive and proactive functions. The reactive function is short-term optimisation of population, operated by denying entry into the city. This function resembles the profit maximising production decisions of firms. It can be implemented in a simple manner without notable political conflicts. Proactive urban governance concerns long-term improvement in the utilisation of the welfare-creating technology of the city. This function is analogous to the reorganisation of a firm's operating practices in utilising the production technology available.

Implementation of this function necessitates considerable effort and is likely to evoke inconvenient equity and other such considerations. Therefore, it is undertaken only under competitive pressure.

A somewhat paradoxical finding is that dynamic efficiency in a system of cities precludes non-stable migration patterns. In stable market situations the competitive dynamics remain partial and sub-optimality of resource allocation results in dynamic sense, because there is no pressure of exit to compel the cities to use proactive urban governance. Stability depends on the initial stage of utilisation of agglomeration economies. If all cities are initially large, meaning that agglomeration diseconomies dominate in all cities, the market solution is stable in nature. This is also true in the case of asymmetric cities, particularly if the larger cities implement reactive optimisation of population. The market situation is clearly non-stable if all cities are initially small, i.e. agglomeration economies dominate everywhere, and both the cities that gain and those that lose population are forced to improve their technical efficiency. Non-stable market situations also exist in an important special case, where small modern cities draw people from larger traditional cities in which agglomeration diseconomies dominate. In these two cases there is a close enough analogy to private firms operating on competitive markets.

Reactive urban governance is a major source of stability, because denying entry into in-migration cities eliminates the pressure of exit in the out-migration cities. If population is optimised with reference to average welfare below the potential capacity of the city, the outcome is socially sub-optimal in the dynamic sense, and in most cases also in the static sense. Optimisation of population is socially beneficial only when connected to long-term growth policy. Optimal population targets should therefore be set only with reference to long-term welfare potentials to ensure the efficiency of the long-term general equilibrium in the city system. It is clear, however, that this function

remains purely theoretical in practice. From the social point of view, optimisation of population must be considered very critically.

The analysis provides a clear-cut principle for national urban policy. On the one hand urban policy must work indirectly to enhance the functioning of the market mechanism. In this respect the pressure of exit must be strengthened and proper use of voice must be encouraged. The pressure of exit can be exerted by social and institutional reforms that enhance the mobility of people and remove distortions from the price mechanism that steers residential choices. To enable better use of voice the autonomous position of the cities as true market agents must be strengthened with particular emphasis on the implementation of proactive instruments of urban governance. Sufficient fiscal and administrative autonomy must be assigned to economically arranged urban areas. The Finnish system of highly autonomous local governments is essentially a solid basis for such development. The urban areas must be made to understand that they are not competing with smaller places and rural areas – they are competing with each other for the welfare increasing migrants, and the competition is becoming more and more international in nature. Therefore, the welfare potential and growth of the largest cities is of particular importance. Short-term optimisation of population must be prohibited and the cities must be motivated to concentrate on proactive urban governance. Harsh structural take-offs must be encouraged, especially in large cities with a traditional and possibly outdated industrial structure, and which seem to be reluctant to accept newcomers.

On the other hand national urban policy has direct functions. The market-oriented mechanism cannot handle important issues such as inter-regional externalities and equity. Traditionally the negative externalities have been of major concern, but nowadays the positive externalities attached e.g. to education, research, health care, and to the formation on social capital in general have received more and more attention. Furthermore, as was indicated above, proactive urban

governance quite evidently raises considerations of vertical equity. Horizontal equity among cities and regions, which has traditionally been a core element in the Finnish welfare state, may also be endangered. All these issues must be dealt with in the framework of fiscal federalism, and clearly national level functions have to be compassed directly by means of urban policy.

The above conclusions are of particular interest in Finland. In the present stage of globalisation and economic integration, the Finnish economy must compete against the circumstances on the European single market. The former comparative advantages, largely attached to the rural sector, have shifted to the new absolute advantages of the urban economy. The important role of agglomeration economies has become evident, which is also reflected in the recent trend towards concentration. Moreover, the undeniable handicap caused by geography and climate serves to stress the fact that spatial factors carry more and more weight in the determination of absolute advantages. In a northern country, the costs of housing and transportation are high compared to its European rivals. Therefore, the traditional element of agglomeration economies, that is the economics of scale and scope in constructing the inter- and intra-urban infrastructure, is still of particular importance. The fact that the development of the Finnish urban sector must be promoted has also been recognised in the recently introduced national urban policy. The policy has been constructed to rest on the intrinsic development of economically determined urban areas, and aims at deeper integration of the communities in those areas. So far there has been insufficient political courage to focus the national urban policy on the issue of city size.

In the light of the analysis above, the issue of city size, measured both in terms of population and geographic area, would appear to be of immense importance to Finnish urban policy and urban governance. Finnish cities vary in size and in their economic, historic, geographic and other such determinants, but nevertheless they are all small and newborn by international standards, and are

scattered over a large country. Even the most rapidly growing urban areas are geographically wide and their population densities remain low.

From this perspective, at least four questions for further research can be formulated. First, there is a need to evaluate the present practices of urban policy and urban governance. Finnish urban policy aims quite ambitiously at developing a wide range of very broadly defined urban areas, and the policy of some key cities seems to be quite negative towards accepting newcomers. It could be argued that a higher degree of concentration would be needed to yield significant agglomeration economies. The assignment of responsibilities between national urban policy and local urban governance must be clarified. Second, the positive feature that the Finnish cities are comparatively autonomous fiscally and administratively does not necessarily concern the functional urban areas. Means of integrating the municipalities in the urban areas should be investigated. Third, it would be worthwhile to explore the true potentials of the larger cities by national and international comparisons e.g. by means of data envelopment analysis. It is a quite reasonable to assume that there must be plenty of room for improvement in their technical efficiency. And fourth, the effects of proactive urban policy should be evaluated. As to competition policy, the emphasis should be on efficiency improvements in the field of local private markets. The effects and means of structural and growth policy should be studied with particular interest.

**References:**


Alonso, W. (1971) The Economics of Urban Size. *Papers and Proceedings of the Regional Science Association*, 26.

Atkinson, A. and J. Stiglitz (1980) *Lectures on Public Economics*. London: McGraw-Hill Book Company Ltd.

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

Boadway, R. and N Bruce (1984) *Welfare Economics*. Oxford: Basil Blackwell Ltd.

Brueckner, J. (1982) A Test for Allocative Efficiency in the Local Public Sector, *Journal of Public Economics*, 19.

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

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

Kajanoja, J. and J. Simpura (eds.) (2000) *Social Capital, Global and Local Perspectives***.** VATT-Publications 29. Helsinki: Government Institute for Economic Research.

Laurila, H. (2000) Urban Governance – kaupunkitalouden tehokas hallinta, *Kunnallistieteellinen aikakauskirja*, 1/2000.

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

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

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

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

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

Prud'homme, R. and C. Lee (1999) Size, Sprawl, Speed and the Efficiency of Cities, *Urban Studies*, vol.35, no. 11.

Tiebout, C. (1956) A Pure Theory of Local Expenditures, *Journal of Political Economy*, 64, October.

Figure 1: Potential and actual welfare in a city

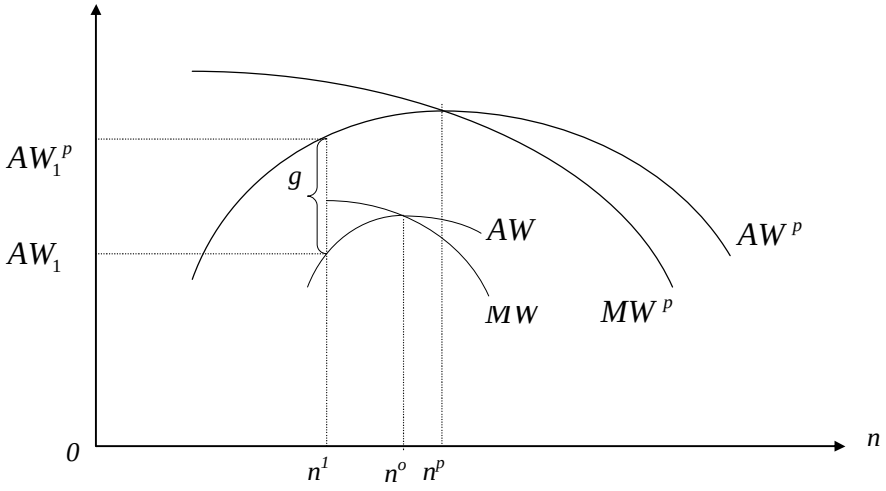


Figure 2: Market equilibrium in large cities

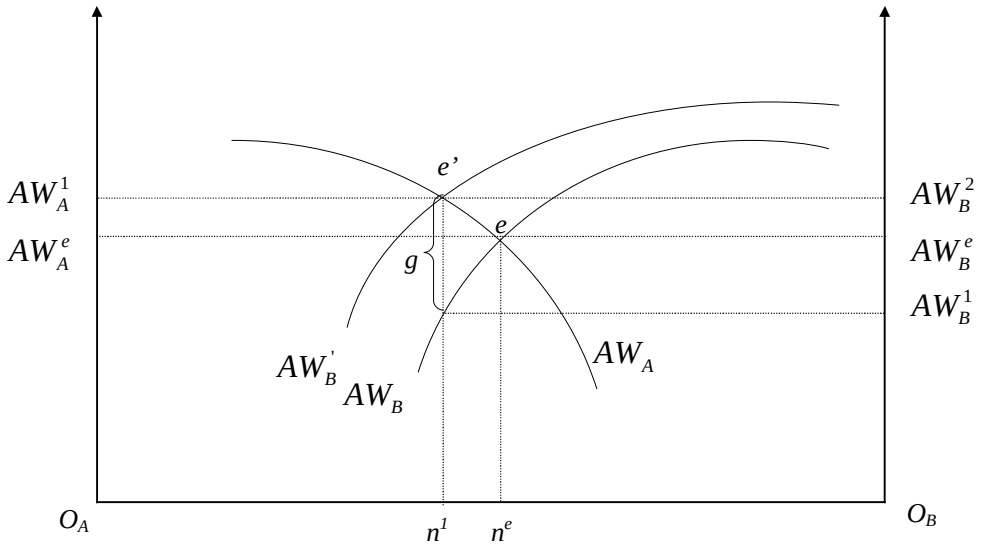


Figure 3: Competitive dynamics in the case of small cities

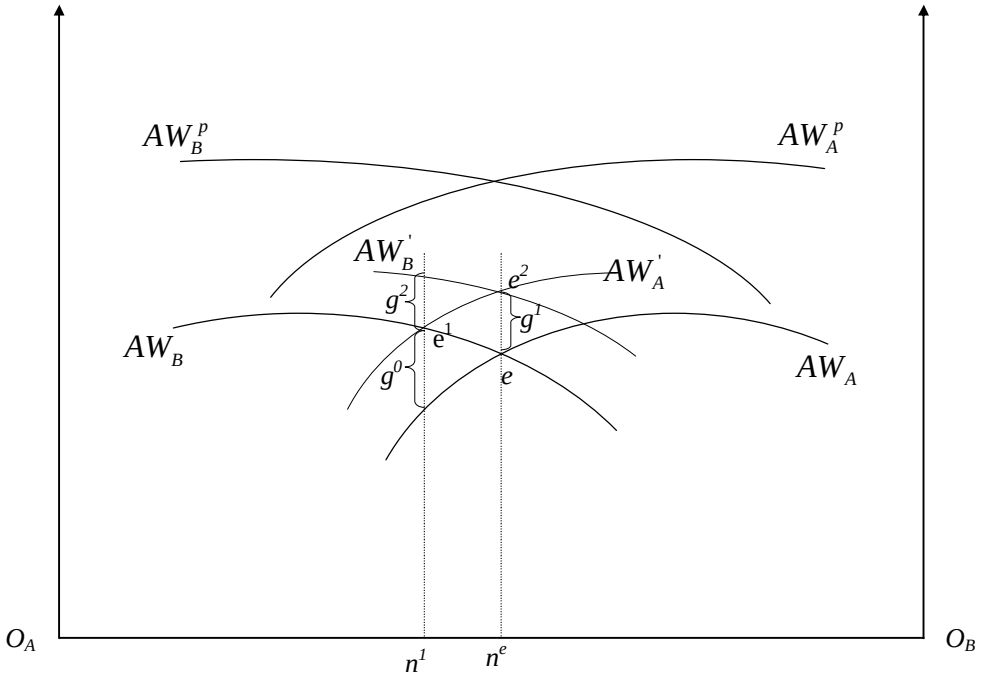
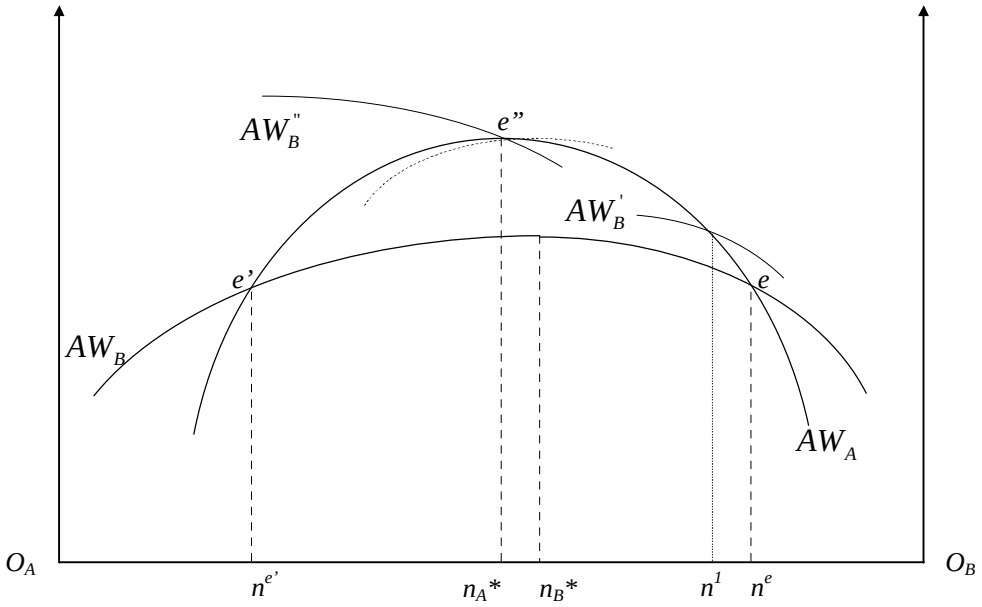


Figure 4: The case of asymmetric welfare potentials



# 5. CONCLUSIONS

## Main contributions of the thesis

The migration of people is an important part of the market mechanism of resource allocation in a spatial context. It ensures the utilisation of all kinds of local economies and diseconomies emerging in the production and consumption of market goods and carries out the allocation of private resources in geographical space. Together with local collective decision-making, migration constitutes a competitive system of local public economies, which yields a market-style allocation in the public sector, too. Precluding the proper conduct of fiscal federalism, the overall outcome is the efficient use of the scarce economic resources in society.

The thesis comprehends the above mentioned properties of migration by taking three approaches on the issue. The analysis starts from **Approach I:** ***The Partial Equilibrium Approach***, which includes Essay #1: *Short-Run Effects of Taxation on Inter-Municipal Migration.* The essay tests the common understanding that, under free mobility, labour income taxation causes people to avoid taxes by migration. Therefore, tax-induced differences in wages should not be sustainable, and the net emigration municipalities, which are under the most obvious pressure to raise their taxes, are doomed to a vicious circle of tax-accelerated emigration.

The novel framework constructed in the essay is a synthesis of the classical labour market model of migration familiar from regional and urban economics, and a model of trade policy familiar from international economics. The analysis shows that migration is not a zero-sum game because it improves social welfare – the in-migration end of the migration flow gains more than the out-migration end loses. Taxation in one municipality causes inter-municipal differences in net real wages in the short term. The distortion suggests that migration and a dead weight loss in social welfare should emerge.

An important finding is that there is an asymmetry in the effects of taxation depending on whether the municipality imposing it is originally an out-migration or an in-migration municipality. An out-migration municipality can levy taxes without affecting migration and welfare in the short term. Only decidedly high taxes provoke such effects. At the other end, even a small tax increase in an in-migration municipality induces emigration and causes both local and economy-wide welfare losses.

The asymmetry in the effects of taxation arises because migration incorporates a vertical segment in the labour supply curve of the out-migration municipality. This migration threshold means that the reservation wage of the remaining worker-residents is much lower than the current market wage. This can be explained by personal factors that determine the workers' nominal wages and the prices of their individual consumption bundles so that both mobility and individually perceived real wages may vary considerably even between people of the same profession. Migration thresholds do not exist in in-migration municipalities. Labour income tax turns the municipality immediately from an in-migration municipality to an out-migration municipality. This is quite reasonable, because taxation evokes a first-time incentive for emigration in the original residents, and a second-time incentive in the recent immigrants. Emigration thus consists of the most mobile of the original residents and of those recent immigrants who have not yet settled firmly.

A general conclusion from the Finnish perspective is that the short-run effects of labour income taxation on migration and welfare are more modest than what is commonly understood. Small taxes in out-migration municipalities have no effects, and even higher taxes have only minor effects when imposed by small and capital-intensive municipalities. At the other end, taxes imposed by big capital intensive in-migration municipalities have no significant effects on migration either. Thus there is

room for inter-municipal tax differences in the Finnish economy without serious threat that taxation would accelerate migration and concentration of population.

**Approach II:** ***The General Equilibrium Approach*** adds to the message of Approach I by considering a much wider variety of economic consequences of migration. This is facilitated by the use of a macroeconomic model, which brings the full set of local markets into the picture. The macroeconomic framework is more suitable for analysing international than inter-municipal issues, but it still yields better understanding of the adjustment to exogenous shocks than do conventional partial equilibrium models. Essays #2 and #3 belong to this category.

Essay #2: *The Macroeconomic Model of Migration* constructs the macroeconomic framework. The main contribution of the essay is to apply the neo-Keynesian macroeconomic model to the analysis of factor migration. The virtue of the model is threefold. First, the model takes a general equilibrium viewpoint on migration instead of the partial equilibrium perspective of the classic labour market model. The full macroeconomic effects in the localities considered can thus be examined. Second, even compared to earlier classical general equilibrium presentations, the present model provides a more detailed analysis of local market adjustment to economy-wide market conditions. Simultaneous analysis of labour markets and goods markets is of particular importance. Third, the model facilitates simultaneous treatment of labour and capital migration. Being quite manageable by graphics, the model yields a major theoretical clarification of the effects of factor migration.

In the model, four types of macroeconomic adjustment paths are specified, namely price adjustment, nominal wage adjustment, opposite adjustment and parallel adjustment. The key finding at this point is that nominal wages may adjust in any direction or even may remain unchanged, but it is for sure

that, in the absence of monetary policy, prices and interest rates must fall due to capital inflow and rise due to capital outflow.

Migration is shown to affect both aggregate supply and aggregate demand conditions in the model. The real effects on the supply side depend on the amount of migration, but the nominal effects, namely those on local prices, nominal wages and interest rates depend on the effects of migration on local aggregate demand. The demand side determines the particular adjustment path, which in turn depends on whether or not the employment effects dominate the wage effects in affecting local consumption demand. If the employment effect substantially dominates the real wage effect, then prices and interest rates move in the same direction and nominal wages in the other. If the dominance is not substantial enough prices, nominal wages and interest rates all adjust in a parallel direction. An important finding is that the local interest rate may rise or fall due to migration depending on the particular adjustment path.

From a theoretical perspective, parallel adjustment seems quite plausible, because other kinds of adjustment require substantial changes in aggregate demand to start them up. The issue remains an empirical one, and depends on the real wage elasticities of labour supply and labour demand on the labour market, and on the real wage difference to be adjusted. Empirically, however, it is arguable that considerable changes do occur – there is convincing evidence that prices rather fall than rise in out-migration localities, and vice versa. This is reasonable, especially from the point of view of housing and real estate prices, which constitute a major part of local consumption expenditures, and which are determined by inelastic supply conditions, at least in the shorter term.

The analysis also illustrates a possible interior solution of spatial allocation of labour in a multi-locality economy. Migration equalises the real wages in the localities, but nominal wages and price

levels are not necessarily equalised. With immobile capital and no trade, big differences in local prices, nominal wages and interest rates may thus exist.

Opposing capital movements mitigate the effects of migration. This is intuitive because the capital inflow increases the domestic productivity of labour and vice versa. An inflow/outflow of capital is a substitute for emigration/immigration, and it can fully offset the need for migration if the impact of capital accumulation is strong enough to make the domestic real wage the same as the market wage in the outer economy. The impact may even be high enough to make the domestic real wage higher/lower than that of the outer economy so that the initial migration pattern is reversed.

Finally, the analysis shows that when both labour and capital are mobile at the same time, it is feasible that there exists an interior market solution of spatial factor allocation between localities. The result holds, however, only in the special case, where the factors are initially drawn in opposite directions by differences in real wages and interest rates, and where the price level is fixed so that the possibility of parallel adjustment is ignored. Theoretically speaking, the conclusion then is that, in the present model without further restrictions and policy considerations, the market mechanism is incapable of ensuring a stable and efficient interior solution of factor allocation.

Essay #3: *Taxes, Transfers and Migration* demonstrates the usefulness of the model of Essay #2 by presenting an analysis of the macroeconomic effects of local public policy on inter-locality migration. A simple tax-transfer programme is introduced in the model and an innovative application of the Adaptive Expectations Hypothesis is used to describe the evolution of people's perceptions about the policy programme over time. The effects of the policy programme are shown to depend on whether the scheme is fair or unfair, which again depends on whether taxes are expected to be fully

compensated or not. Even an actually fair programme may be anticipated to be sub-fair in the short term, during the AEH type adaptation path towards long-term equilibrium.

A fair and correctly perceived policy programme does not affect migration. This is because a fair system leaves the local real wage unchanged and thus has no effects on inter-locality welfare comparisons. The free migration equilibrium remains efficient. The result holds around a long-term market solution. Sub/super-fair policy programmes have short-term effects that encourage emigration/immigration. The effect is the stronger the farther away the short-term equilibrium is from the long-term equilibrium. As to the effects of a sub-fair programme, a pure tax system with no (anticipated) repayments quite intuitively has the strongest effect on emigration, but even partial(ly anticipated) repayments from the system undermine the motives to emigrate. The short-term migration equilibrium is distorted because the workers' real wage deviates from that encountered by the firms.

The short-term effects of sub/super-fair programmes are eventually mitigated in the longer term, provided that the tax revenue is not wasted, there are no redistributive policy goals, and/or there are neither positive nor negative externalities involved in the programme. Under these circumstances an AEH type adjustment in the supply of labour leads to convergence towards the efficient long-term equilibrium.

The results of the analysis of the tax-transfer programme can be generalised to concern the provision of local public goods and services by interpreting the tax wedge in the local labour market as the share of public production, and the corresponding price wedge in the local goods market as the share of free public goods, or transfers-in-kind, in the local economy. The analysis describes what happens to the local economy when part of people's consumption needs is provided publicly free of charge

against tax payment. The result is that, under the conditions specified above, the economy remains unchanged in real terms. Thus, there may well be policy differences between localities, at least as far as there are no inter-locality spill-overs. Free migration thus does not automatically lead to equalisation of policy programmes between localities.

The contribution of **Approach III: *The Club Theoretic Approach***, is twofold. First, the approach also allows the endogenous variables in peoples' utility maximisation problem to be locally dependent, thus bringing the issue of technical externalities into the analysis. Migration also internalises these externalities in private choices. Second, the model merges the concept of local market area with the concept of local public economy. The residents of a locality are not only able to migrate between localities, but they can also collectively affect the local conditions of their welfare. This concerns both the provision of local public goods and the functioning of the local private markets. Essays #4, #5 and #6 belong to this category.

Essay #4: *The City as a Club* constructs the club theoretic model based on the interdependence between city size and welfare in the city. The interdependence is explained by agglomeration economies and diseconomies linked to city size, which are internalised into people's benefits and costs of city life. People respond to possible inter-city differences in welfare by migrating.

The analysis shows that free migration of people alone cannot constitute an efficient market mechanism. The city itself must act as a market agent and choose its optimal population. In operating this function, there are two competing criteria of optimality, the within-club viewpoint that aims to maximise individual welfare of the residents of the city, and the total economy viewpoint that aims to maximise the welfare of the city as a whole. These criteria yield different solutions: in terms of efficiency the total economy viewpoint is superior to the within-club viewpoint. The results

show that optimisation of population is a vital function of local governance and suggest that the city should take the total economy viewpoint and maximise the total welfare experienced in the city even though it lowers the individually perceived welfare levels. It might, however, be argued that the within-club viewpoint is more relevant in practice than the total economy viewpoint. This is due to four reasons.

First, the total-economy viewpoint necessitates calculations of total welfare and/or marginal benefits and costs. These concepts are far more abstract than average welfare, which is observed by the citizen themselves in everyday experience and revealed by polls and by systematic migration patterns. Second, the total-economy viewpoint violates both the consistence and coherence of decision-making. It is inconsistent to assume that people make their private migration decisions with respect to one optimisation problem, and the simultaneous collective decisions with respect to another problem. The total economy viewpoint is incoherent with the key feature of club theory, namely that any market mechanism should be based on private optimisation. The formation of clubs should be driven by individually reasoned choices both in private and collective contexts. Third, it can be argued that the within-club viewpoint is implicit in the system of representative democracy. Given that people are consistent in their utility maximisation, they will vote for those decisions/representatives that promote their personal welfare. Candidates who campaign for the total economy viewpoint would simply not be elected. Fourth, if the local government should disregard the preferences of the voters by taking a total economy viewpoint, the voters would still have the option to vote with their feet and so nullify the social planner's optimisation. The total economy type solution is thus not necessarily stable.

Essay #5: *Migration, Policy and the Formation of Cities* uses the model of Essay #4 to examine the working of the spatial market mechanism. The analysis shows that the migration of people alone can

yield efficiency only in the special case of homogenous people and large and homogenous cities. But even in this case, a Pareto superior outcome can be achieved in the longer run if local governments restrict mobility from the within-club viewpoint. The long-run equilibrium necessitates that the number of cities increases so that optimal utilisation of agglomeration economies is reached in all cities. In the case of small cities, migration will lead to corner solutions without local optimisation of population. A Pareto efficient optimum necessitates that the growing cities optimize on their population, and that the number of cities falls as an endogenous variable in the long term. Total population can also be taken as a variable, but if it is allowed to decrease the welfare implications become more complicated unless a global perspective is adopted.

The case of asymmetric cities and homogenous people states that a stable and efficient market solution is not achievable at all. In this case not even local optimisation of population is sufficient to sustain efficiency, and centralized policy is necessary. Two options of centralized policy are studied, namely administrative regulation in the form of quantity rationing, and economic measures in the form of lump sum transfers.

Administrative regulation can be used to maintain the efficient allocation of people in spite of the existing welfare differentials between cities. Such measures may be difficult to implement because they necessitate constraints on mobility. Transfers between the cities constitute a more applicable policy measure. Lump sum transfers from richer to poorer cities can be used to equalise the welfare differentials.

The social welfare effects of administrative and economic policy measures are the same but, because of the welfare equalising 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 maximal

individual welfare is achieved 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 fewer people live in the more prosperous cities than in the regulation option. The finding is important because it not only shows that the effects differ but also reveals that effective transfer policy necessitates endogenous determination of the number of cities. Lump sum transfers actually necessitate reallocation of people between cities compared to administrative regulation, and also compared to a stable market solution based purely on migration. Therefore, lump sum transfers rather accelerate than stabilise migration, which is a result that contradicts the conventional wisdom that lump sum transfers should be neutral in their effects on individual choices.

The Tiebout hypothesis states that formation of an efficient set of heterogeneous cities with homogeneous populations within them is feasible. If all these cities optimize on their population and no integer problem exists, the outcome is a Pareto efficient long-term solution. Of course, in this case it must also be precluded that there are no inter-city externalities or other issues that would be of national-level interest.

Essay #6 *Urban Governance, Competition and Welfare* yields strong policy conclusions. The essay complements the message of the Tiebout model of spatial resource allocation by investigating the preconditions of competitive dynamics in a system of cities. Local welfare is maximised by urban governance (voice) under the market pressure created by migration (exit). The dynamic efficiency of the spatial system rests on the interplay of exit and voice. The incentives for dynamic improvements depend on the competitive nature of the system.

Urban governance is divided into reactive and proactive functions. The reactive function is short-term optimisation of population, operated by denying entry to the city. This function resembles the

profit maximising production decisions of firms. It can be implemented in a simple manner without notable political conflicts. Proactive urban governance concerns long-term improvement in the utilisation of the welfare-creating technology of the city. This function is analogous to the reorganisation of a firm's operative practices in utilising the available production technology. The implementation of this function entails considerable effort and evokes inconvenient equity and other such considerations. Therefore, it is undertaken only under competitive pressure.

An important finding is that dynamic efficiency in a system of cities precludes non-stable migration patterns. In stable market situations the competitive dynamics remain partial, because there is no exit pressure to compel the cities to use proactive urban governance. The market solution is stable when all cities are initially large, or when cities are of asymmetric size and the larger cities implement reactive optimisation of population. The market situation is clearly non-stable if the cities are initially small and both in- and out-migration cities are forced to make improvements in their technical efficiency. Such situations also exist, when small modern cities draw people from larger traditional cities, in which agglomeration diseconomies dominate. In non-stable cases there is an analogy to private firms operating on competitive markets.

Reactive urban governance causes stability, because closing entrance into in-migration cities eliminates the exit pressure in the out-migration cities. If population is optimised below the potential capacity of the city, the outcome is socially sub-optimal in the dynamic sense and often also in the static sense. Optimisation of population is socially beneficial only when connected to long-term growth policy. Optimal population targets should therefore be set with reference to long-term welfare potentials. It is clear, however, that this function remains theoretical in practice. An important extension to the findings of Essays #4 and #5 is that optimisation of population must actually be assessed very critically.

National urban policy has indirect and direct functions. The indirect function is about enhancing the working of the market mechanism by strengthening the exit pressure and encouraging proper use of voice. Sufficient fiscal and administrative autonomy must be assigned to economically defined urban areas. Short-term optimisation of population must be prohibited and the cities must be motivated to concentrate on proactive urban governance. The direct functions mostly concern inter-regional externalities and equity, which can be best dealt with in the framework of fiscal federalism.

### Final remarks

The thesis provides an original insight into the migration of people and sets up interesting frameworks for further use both theoretically and empirically. Simple policy experiments show that original contributions can be derived. Some guidelines for further work can be noted.

The partial equilibrium framework of Approach I would clearly benefit from more sophisticated treatment of the labour markets. The existence, rationale and effects of the migration threshold should be studied more closely. In this respect, theories of segregated labour markets, efficiency wages etc. would yield better intuition theoretically and serve empirical work better, too. The effects of the firms' possibilities for rent-seeking under the migration threshold should be analysed. The tax policy experiment of Essay #1 is simplistic, and a fuller definition of taxation with redistributive elements might yield interesting results.

The macroeconomic model of Approach II is very promising and rich in detail, not all of which was considered in the thesis. The framework of Essay #2 raises many interesting questions for theoretical and empirical work, although not all of them can be assessed on a municipal level. The model allows for a wide range of policy experiments, including both micro and macroeconomic policies. The

policy effects could also be considered in different policy regimes. The expectations, adaptive and/or rational expectations, could also be modelled more explicitly. The simplistic tax-transfer programme of Essay #3 could be elaborated to include redistribution and progression. A valuable amendment of the model might be to consider the policy-induced non-linearity of the real wage schedules. One promising line of theoretical research would be to challenge the conventional (one-sector and two-sector) models of regional factor allocation.

Approach III yields interesting research questions particularly for empirical study. An important question is the true existence and utilisation of agglomeration economies. Furthermore, the welfare potentials of larger cities could be evaluated by national and international comparisons (e.g. by means of data envelopment analysis). It is quite a reasonable assumption that there must be plenty of room for improvement in the technical efficiency of the cities. The measurement of the effects of inter-city competition would be a more challenging research object, and so would be also an attempt to add something to the literature on testing the Tiebout hypothesis.