

29/02/20:

Why some places are left-behind: urban adjustment to trade and policy shocks

Anthony J. Venables*

University of Oxford and Monash University.

Abstract

Economic adjustment to trade and policy shocks is hampered by the fact that some sectors tend to cluster, so are hard to initiate in new places. This can give rise to persistent spatial disparities between cities within a country. The paper sets out a two-sector model in which cities divide into those producing tradable goods or services subject to agglomeration economies, and those only producing non-tradables for the national market. If import competition destroys some established tradable sectors, then affected cities fail to attract new tradable activities and switch to just produce non-tradables. Full employment is maintained (we assume perfect markets and price flexibility) but disparities between the two types of cities are increased. All non-tradable cities experience real income loss, while remaining tradable cities boom. The main beneficiaries are land-owners in remaining tradable cities, but there may be aggregate loss as the country ends up with too many cities producing non-tradables, and too few with internationally competitive activities. Fiscal policy has opposite effects in the two types of cities, with fiscal contraction causing decline in cities producing non-tradables, increasing activity in cities producing tradable goods, widening spatial disparities, and in the process increasing the share of rent in the economy.

Keywords: Urban economics, divergence, rebalancing, lagging regions, de-industrialisation, fiscal policy.

JEL classification: F12, F60, R11, R12.

* Thanks to Paul Krugman, David Soskice and David Vines for helpful comments.

Author's address:

A. J. Venables,
Department of Economics,
Manor Road,
Oxford OX1 3UQ, UK
tony.venables@economics.ox.ac.uk

1. Introduction

The premise of this paper is that some jobs are difficult to move and hard to establish in new places. Some sectors cluster -- finance in London, tech, creative sectors, motor vehicles in other centres -- and firms are unwilling to move out of their cluster. How, under these circumstances, does the economy respond to shocks such as changes in competitiveness and macro-economic policy? If sectoral clustering translates into spatial heterogeneity then the impact of shocks varies across places. Do adjustment process tend to equalise these impacts or to amplify spatial differences, leaving some places ‘left-behind’? Who – and where – are the gainers and losers from macro-economic policy?

The paper is analytical, setting out a simple model to answer these questions. Building on the single feature of agglomeration economies in some productive sectors, we establish a number of propositions. There is spatial heterogeneity in production structure, wages and productivity. Adjustment to trade shocks can result in outcomes that, while maintaining full employment, leave a country with too few places that have internationally competitive sectors (producing tradable goods or services), and too many producing just for the domestic market (non-tradables). The latter have lower productivity, nominal wages, population, house prices and land values. Fiscal policy has spatially differentiated impacts, with some cities contracting, others booming, and landowners in booming cities benefiting relative to workers and landowners elsewhere.

The model we develop has a number of ingredients. The first is that it is spatial, containing different places. We label these ‘cities’, and draw on standard techniques of urban economics to describe them. The second is that there must be a reason why jobs are hard to relocate, and here too we draw on urban economics. Agglomeration economies are a central feature of cities and often occur within particular sectors (sometimes referred to as localisation economies). They take the form of increasing returns to scale that arise between firms (or between firms and workers) in a particular place and type of activity. They are driven by thick market effects arising from proximity of specialist workers, suppliers, and customers, and by knowledge spillovers and intense competition. These are reciprocal externalities that increase the productivity of firms that can access these advantages.¹ Two consequences of agglomeration economies are central to this paper. One is that – since access is largely a function of proximity – spatial clusters of activity develop. The other is the ‘first-mover’ problem. No firm wants to move out of a cluster, as it would forego the productivity advantages that it provides; a coordinated move of the entire cluster might be profitable, but coordination failure creates locational inertia.

¹ For theory and evidence on agglomeration economies see Duranton and Puga (2004), Rosenthal and Strange (2004) and Combes and Gobillon (2015).

The third ingredient is that there are sectoral differences in the importance of agglomeration economies. Some sectors exhibit a strong tendency to cluster. In others agglomeration economies are relatively weak and location decisions are dominated by the advantages of being close to final customers (or, in other sectors, close to resource endowments). We capture these differences in a stark way, by assuming just two types of sector. One we refer to as ‘tradables’; in this sector agglomeration economies create increasing returns and, since output is internationally tradable, prices are set on world markets and the scale of production is not limited by the size of the market. Examples are finance, much high-technology, some creative sectors, and also more traditional sectors such as garments. The other is ‘non-tradables’, meaning goods that are sold on an integrated national market but are not tradable internationally. In this sector we assume that there are no agglomeration benefits and diminishing returns arise, if not because of diminishing returns in production, because the output price is determined in the national market so is reduced by increases in output. Examples are warehousing, much of the food industry, customer services, back-office operations, and government services.

Working with just these two types of sector ignores products that are traded neither internationally or inter-regionally – for example haircuts and restaurant; these could be added without changing the qualitative results of the model, and are ignored just to avoid clutter. A more general model might include a continuum of degrees of tradability, and a continuum of degrees of agglomeration. We look at just two types -- tradable with agglomeration, and non-tradable without – in order to derive the main messages in the most direct way possible.

The final ingredients are workers and households. Worker/ households are free to choose where to live and work, both between and within cities, facing no migration costs or relocation frictions whatsoever.² Furthermore, all wages and prices are perfectly flexible. The only friction in the model is that arising from agglomeration economies in the tradable sector, and the consequent first-mover problem.

What do we find? First, agglomeration forces mean that there will be a dichotomous city structure, with some cities specialising in tradable goods, and others in non-tradables. We call the former type-T cities, and the latter type-N cities. Second, the friction created by agglomeration induces a non-uniqueness of equilibrium. For given parameter values there is a set of equilibria, and history (or other forces outside the model) determines the point in this set at which the economy finds itself. Different points in the set – i.e. different numbers of type-T cities relative to type-N – vary in city size, nominal wages, house prices, and land rents. Aggregate welfare also varies across this set of equilibria, so it is possible to end up in a

² They must however live and work in the same city; there is commuting within cities, but not between them.

situation with ‘too many’ type-N cities relative to type-T. Type-N cities are stuck in a low-level trap, potentially dragging down aggregate real income.

Perfect labour mobility and price flexibility mean that full employment is maintained. However, agglomeration means that the impact of shocks may be quite different across different cities. For example, a negative trade shock might destroy the traditional tradable activity of a type-T city. Because of the first-mover problem, the market mechanism fails to replace this with another tradable activity, so the city defaults to being type-N. This affects not just that city, but also has an adverse effect on all other type-N cities via increased supply and a lower relative price of non-tradables. There is then a growing set of type-N cities each experiencing population loss, low rents and house prices, and low nominal wages (although real incomes are equalised by spatial variation in land and house prices). Fiscal policy will have different effects on cities of different types and we show that contractionary fiscal policy will increase regional disparities as declining type-N cities exist alongside booming type-T cities.

The rest of the paper develops the simplest possible model to make these statements precise. Labour markets in the model are ‘perfect’, and the model shows how a single feature – agglomeration economies in tradable sectors – can create quite divergent urban performance. The structure provides an analytical framework for thinking about economies in which booming cities coexist with others in decline, and the policy challenges faced in these economies.³ Some generalisations and extensions of the model are discussed in section 6 of the paper and section 7 outlines alternative policy responses.

2. The urban dichotomy: a model

A small open economy contains a fixed number of cities, M , and is endowed with a single factor of production, L units of labour, that is perfectly mobile between cities. Two forces determine the distribution of population between cities. One is that cities offer jobs, and the other is that larger cities are more expensive because of commuting costs and land prices.

Jobs are either in the production of tradable goods and services, or non-tradables. Tradables have their price, p_T , fixed on world markets. They can be thought of as a number of different goods or sectors, but all are symmetric and price-taking. Their production involves city specific agglomeration economies (localisation economies), so that productivity in tradable production in a particular city is increasing in the number of workers employed in tradables in the city. Denoting this employment by L_T , output per worker is $q(L_T)$, with $q'(L_T) \geq 0$. Since

³ See Storper (2018) for a discussion of literature on regional economic polarisation, and Haldane (2019) for the U.K. position.

labour is the only input to production and producers make zero abnormal profits, the wage is the value of output per worker,

$$w_T = p_T q(L_T). \quad (1)$$

This relationship is central to what follows, and means that the wage offered by tradable employment in a city is increasing in the level of that employment. We shall choose units such that, if the sector is not present in the city, the productivity of any small firm that starts tradable production is unity, $q(0) = 1$.

Non-tradables are freely traded within the country but not internationally; they can be thought of as some food sectors, customer services, warehousing, logistics, government services, and perhaps the back-room jobs of the financial and insurance sectors. They are produced under constant returns to scale and their price p_N is set on the national market. Choosing units such that each worker produces one unit of output, this price is equal to the wage offered in the sector, $w_N = p_N$; henceforth we will use this equality, and sometimes refer to the price of non-tradables as w_N .

The presence of agglomeration economies in the tradable sector means that each city will specialise in either tradable production or non-tradable. (A city could only do both if it offered the same wage in each sector, but this is unstable, as a small increase in tradable employment would increase productivity and w_T). Cities therefore divide between those specialised in tradables (type-T cities), the number of which we denote M_T , and those just producing non-tradables (type-N), of which there are $M - M_T$. Cities of each type are symmetric so that all type-T cities have the same population, L_T , and all type-N are of size L_N . Full employment is therefore

$$L = M_T L_T + (M - M_T) L_N. \quad (2)$$

As noted above, the price of non-tradables, w_N , is set on the national market and equates supply and demand. It therefore solves the equation

$$(M - M_T) L_N w_N = \theta [M_T L_T w_T + (M - M_T) L_N w_N] + G. \quad (3)$$

The left hand side is the value of supply of non-tradables and the right hand side is the value of demand, where the term in square brackets is total income generated in the cities, fraction θ of which is spent on non-tradables, the remainder going on tradables. G is net government expenditure on non-tradables. This spending pattern implies that the price index in the country (the same in all cities since all goods are freely traded within the country) is⁴

⁴ P is the price index of a composite of tradables and non-tradables. It does not include the cost of housing and commuting, given below.

$$P = p_T^{1-\theta} w_N^\theta. \quad (4)$$

These relationships determine the wages offered in cities of each type. It remains to determine their population. Workers are perfectly mobile and choose where to live, deriving utility from living in a city of type- i given by $u_i = (w_i - bPL_i)/P$, $i = N, T$. This is the real wage net of ‘urban costs’, the expression bPL_i . These costs can be thought of as commuting costs and rent, and will be discussed further below. They increase with city population L_i at rate b , have unit cost P , and have to be paid by each household.⁵ Utility is then remaining income, deflated by the price index P . Labour mobility equalises utility across all cities, which occurs when the population is divided such that

$$w_T - bPL_T = w_N - bPL_N. \quad (5)$$

Thus, larger cities have to pay a higher nominal wage in order to offset the higher costs of rent and commuting.

Equilibria: We analyse the model diagrammatically by expressing wages offered by each type of city, w_T, w_N , as a function of the proportion of cities that are type-T, M_T . This requires that the five simultaneous equations above are solved for variables $\{w_T, w_N, L_T, L_N, P\}$, as functions of parameters and the number of type-T cities, M_T ; since the equations are non-linear they are solved numerically (see appendix).

Fig. 1 shows the two wage relationships, with the proportion of cities of type-T on the horizontal axis and wages on the vertical. The curve w_N gives the wage that is paid in type-N cities and it slopes upwards because, moving to the right, there are fewer type-N cities, and hence less supply of type-N goods that therefore fetch a higher price and support a higher wage. The curve w_T is the corresponding curve for type-T cities. The price of type-T goods is fixed on world markets but, moving to the right, there are more type-T cities, each of which is therefore smaller. Because each city is smaller it has lower agglomeration economies and lower productivity, this dragging down wages.

To find an equilibrium in this model we have to ask three questions. Is it worthwhile for a worker to move from one city to another? Is it worthwhile for a non-tradable firm to move from a type-N city to a type-T city? And is it worthwhile for a tradable firm to move from a type-T city to a type-N city? If all three answers are negative, we have an equilibrium.

Looking first at workers, there is perfect mobility, but it does not follow that nominal wages are the same in cities of both types. Cities of different sizes have different costs of living so, if eqn. (5) is satisfied, there is no incentive for any worker to move. We have already built this

⁵ Commuting costs are incurred in units of this composite good, and rent is spent entirely on this good. The assumption that these costs are linear in L_i is discussed in section 6.

into the solution of the model so the figure, as constructed, has the gap between the wage curves satisfying eqn. (5), $w_T - w_N = bP(L_T - L_N)$.

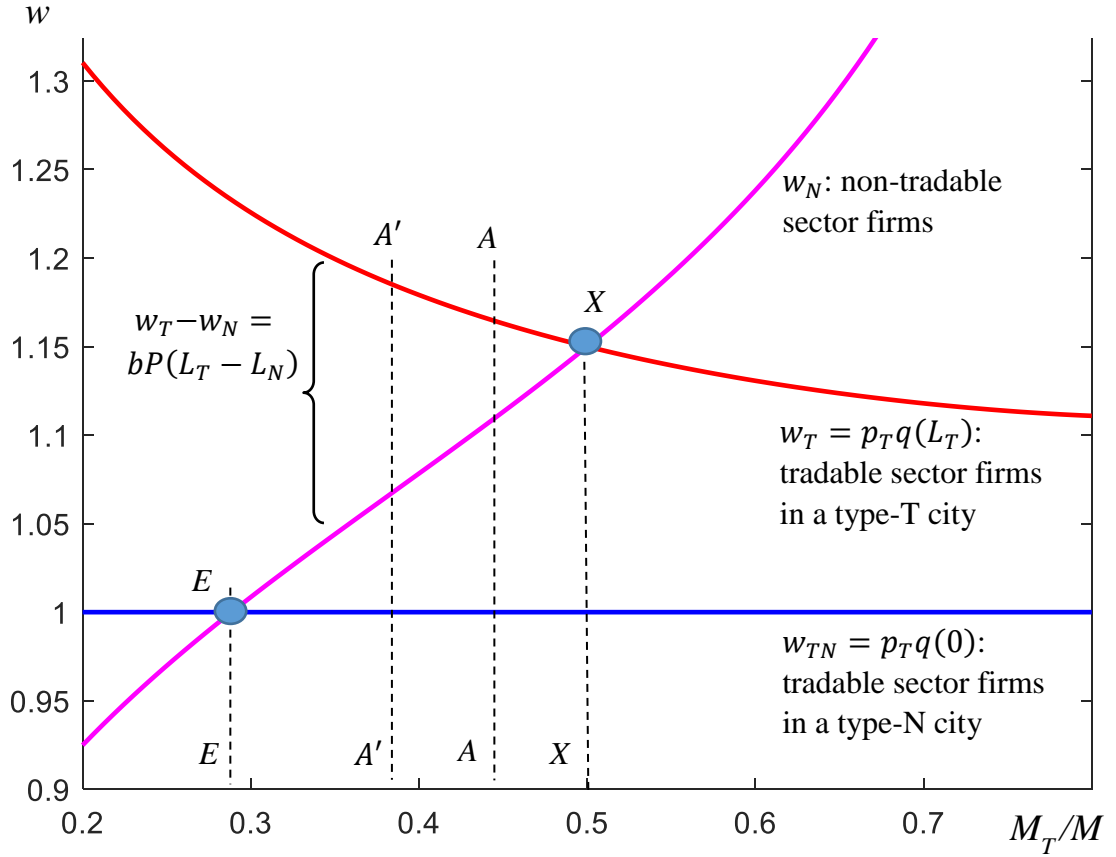
Is it worthwhile for a non-tradable firm to move to a type-T city? This depends only on the wages in the two cities, and it is worthwhile if wages in type-T cities are lower than in type-N, $w_T < w_N$. This means that no point to the right of the intersection at X is an equilibrium. Essentially, to the right of X there are so few type-N cities that the price of N-goods is high enough for it to be profitable to set up a non-tradable firm even in a type-T city.

Finally, is it worthwhile for a tradable firm to move from a type-T to a type-N city? The alternatives are to stay in a type-T city, where wages are equal to the value of output per worker, $w_T = p_T q(L_T)$, or move to type-N where the value of output per worker would be $p_T q(0)$. The move is profitable if this is greater than the type-N city wage, w_N . On fig. 1, this is points to the left of the E , the intersection of w_N with the line $w_{TN} = p_T q(0)$ (where w_{TN} stands for the wage at which a tradable sector firm would break even in a type-N city). To the left of E there are so many type-N cities that the price and wage w_N is low enough for a tradable sector firm to produce profitably even though, moving to a type-N city it forgoes productivity benefit $q(L_T) - q(0)$.

From these arguments it follows that any value of M_T/M in the interval $[E, X]$ is an equilibrium. For example, if the proportion of cities that are type-T is at the vertical line AA then tradable cities pay higher wages, but this is consistent with labour mobility as these cities are larger and have a higher ‘urban costs’ of commuting and rent. Non-tradable firms do not want to move to such high wage cities ($w_T > w_N$), and the wage gap is not large enough to compensate tradable sector firms for foregoing the agglomeration benefits of staying in their high-cost and high-productivity type-T cities ($w_N > w_{TN}$). The fundamental reason is the coordination failure – no firm wants to be the first to establish tradable sector production in a place with no existing production in the sector.⁶ Notice also that the coordination failure is one-sided, given the assumption of increasing returns in tradables, but not in non-tradables. It follows that at points in the interval (E, X) relative wages satisfy $w_T > w_N$, and hence type-T cities are larger than type-N, $L_T > L_N$.

⁶ In the absence of coordination failure (e.g. if a large developer can create the coordinated establishment of a larger number of firms) the equilibrium set would shrink to point X on Figure 1. At this point cities of each type are the same size.

Figure 1: Wages paid by each sector in each type of city



This framework shows how cities of different types coexist, and how the division of cities between the two types is not uniquely determined. How does this system react to shocks? As these occur, who gains and who loses? Analysis will involve shifting some of the curves on Fig. 1. In section 3 we look at a trade shock the impact of which is to shift the line AA (the initial value of M_T). In section 5 the direct effect of fiscal policy is to shift both the w_T and w_N curves, thereby moving the equilibrium set, $[E, X]$. In each case the nature of the economy's response depends on whether or not AA remains in the interior of this equilibrium set.

There is one further point by way of set-up. The 'urban costs' borne by each worker are a combination of rent and commuting costs. In the standard urban model this division depends on residential location within the city. For a worker at the city edge they are entirely commuting costs while at the city centre they are entirely rent.⁷ If the city is linear and commuting costs are linear in distance then the commuting cost paid by the marginal worker

⁷ Rents adjust until workers are indifferent about where in the city they live. For detailed exposition of the Alonso-Muth-Mills model of urban land-use see Duranton and Puga (2015).

(living at the city edge) is bPL_i , ($i = T, N$), and the total of commuting costs and rent city-wide is bPL_i^2 . Total commuting costs are half this, the remainder being rent, $R_i = bPL_i^2/2$. Rents are therefore increasing and convex in city size and, if parameter b takes the same value in all cities, then both total real rents and total real commuting costs are minimised when all cities are the same size. To the left of point X type-T cities are larger than type-N, so it follows that a change that expands type-T cities will increase real land rent (see appendix). The effect on total real income (the sum of the real income of workers plus land rents) is ambiguous, depending on the extent of returns to scale in tradable sectors.

3. Import competition

The impact of international competition and technical change has often fallen on particular sectors such as textiles, shipbuilding, or extractives, and the cities and towns where they are concentrated. How does the economy adjust to such exogenous shocks? The answer comes from Fig. 1.

Taking AA as the initial position, we suppose that exogenous change (import competition) causes $A - A'$ of these cities to lose their tradable sectors, i.e. the leftwards shift to $A'A'$ illustrated in Fig. 1. Initially each of these cities was type-T city paying wages w_T , and the loss of their tradable sectors means that the nominal wage falls until it hits w_N , at which point non-tradable sector firms find it profitable to set up in the city. Each of these cities therefore switches from being type-T to type-N. The economy is flexible enough for new jobs to be created but the first-mover problem means that these are not in new tradable activities, but are instead in non-tradables. Associated with this nominal wage reduction there is out-migration from each affected city (since $L_T > L_N$) and hence house prices and land rents fall until household utility is equalised across city types (eqn. 5).

There are further implications throughout the economy. As affected cities switch from type-T to type-N, so the equilibrium moves to the left on Fig. 1. There is increased supply of sector-N goods so their price and wages w_N fall in *all* type-N cities, leading to out-migration from all type-N. Workers move to type-T cities, bidding up land-rents and house prices in these cities and, if there are agglomeration benefits at the margin, raising productivity. The final resting place is as follows:

- Affected cities switch from type-T to type-N.
- All type-N cities experience out migration and have lower nominal wage than before.
- The share of the population in type-N cities increases.
- Remaining type-T cities have higher nominal wages and are larger than before.
- The share of land-rent in national income increases.

Notice that in this process real wages are equalised across cities (following directly from the assumption of perfect labour mobility) although nominal wages diverge. The principal losers

are people who own land in type-N cities – all type-N cities, not just those that suffer the direct shock, as the effect is shared via a fall in the price of sector-N goods. The gainers are landowners in the type-T cities. These cities boom as there is an influx of labour, possibly enhancing agglomeration economies and raising productivity. Rents are bid up, by how much depending on the supply of land and housing in the city. Since type-T cities are larger than type-N, moving workers to these larger cities raises the share of rent in the economy.

What about aggregate welfare? If agglomeration economies in each city are, at the margin, exhausted, (i.e. $q(L_T)$ is constant, although greater than $q(0)$) then further changes in L_T have no effect on productivity. In this case aggregate welfare is maximised at the intersection point X , where $w_T = w_N$. Moving away from this point – moving to the left in the interior of $[E, X]$ – then reduces aggregate welfare and the economy is made worse off. If however there are further agglomeration economies at the margin then, as the number of type-T cities falls and the size of those remaining increases, productivity in these cities increases, an effect that could be large enough to generate aggregate welfare gain.

This arguments apply if the equilibrium remains within the range $[E, X]$ on Fig. 1. Reaching the edge of this range, the adjustment mechanism switches to one where new tradable activities replace those that are lost, as will be explored in following sections. The message from this section is that the economy adjusts to trade shocks and maintains full employment, but with a falling share of tradables production and an increasing dichotomy between city types, with type-N shrinking and type-T booming.

4. Export booms

Some tradable sectors receive positive shocks due to growing world demand for their output or technical change that raises their productivity. The effect of a positive shock of this type depends on whether or not it affects all tradable sectors. In the unlikely event that all T-sectors are affected, it spills over positively throughout the city system. The direct effect is to raise wages in tradable sectors and cities (this shifting the w_T curve on Fig. 1 upwards). This sets in train two other forces. One is that this additional income generates spending on sector-N goods so their price, and hence the wage curve for type-N cities, w_N , also shifts up. The increase in w_N is typically less than the increase in w_T , so there is migration from type-N cities to type-T until real wages have once again been equalised. Overall, this is a story of complementarity as booming tradable goods sectors also benefit the rest of the economy.

Tradable goods sectors are, in reality, heterogeneous. There are ‘new’ sectors in technology, creative sectors, and finance, and ‘old’ ones, such as many manufacturing sectors. What if the positive shock is restricted to the ‘new’ sectors? These sectors will expand, raising both the nominal wage and employment in affected cities. This wage increase is however a *negative* shock for old-sector tradable cities. They are competing for labour and the wage has gone up,

undermining their competitiveness. If the survival of some of these sectors is marginal then the wage increase puts them out of business, and further effects are as described in the preceding section. Old-sector type-T cities switch to type-N production, the supply of N-goods increases, reducing w_N , and leading to contraction of all type-N cities. The net effect is that, while tradable cities that receive the positive shock boom, other tradable cities and all non-tradable cities are negatively affected. And as before, much of the effect ends up in changes in rents, with these booming in the growing cities and falling everywhere else.

The intuition behind this result is best understood in terms of the ‘Dutch disease’ effect of resource rich economies. A booming sector will draw labour from other sectors and, in the present setting, other cities. An increase in the world price of one set of tradable sectors is technically identical to a reduction in the price of the other set. In each case exit of the *relatively* badly affected type-T cities occurs, increasing the number of type-N cities with adverse effects on all of them.⁸

5. Fiscal policy

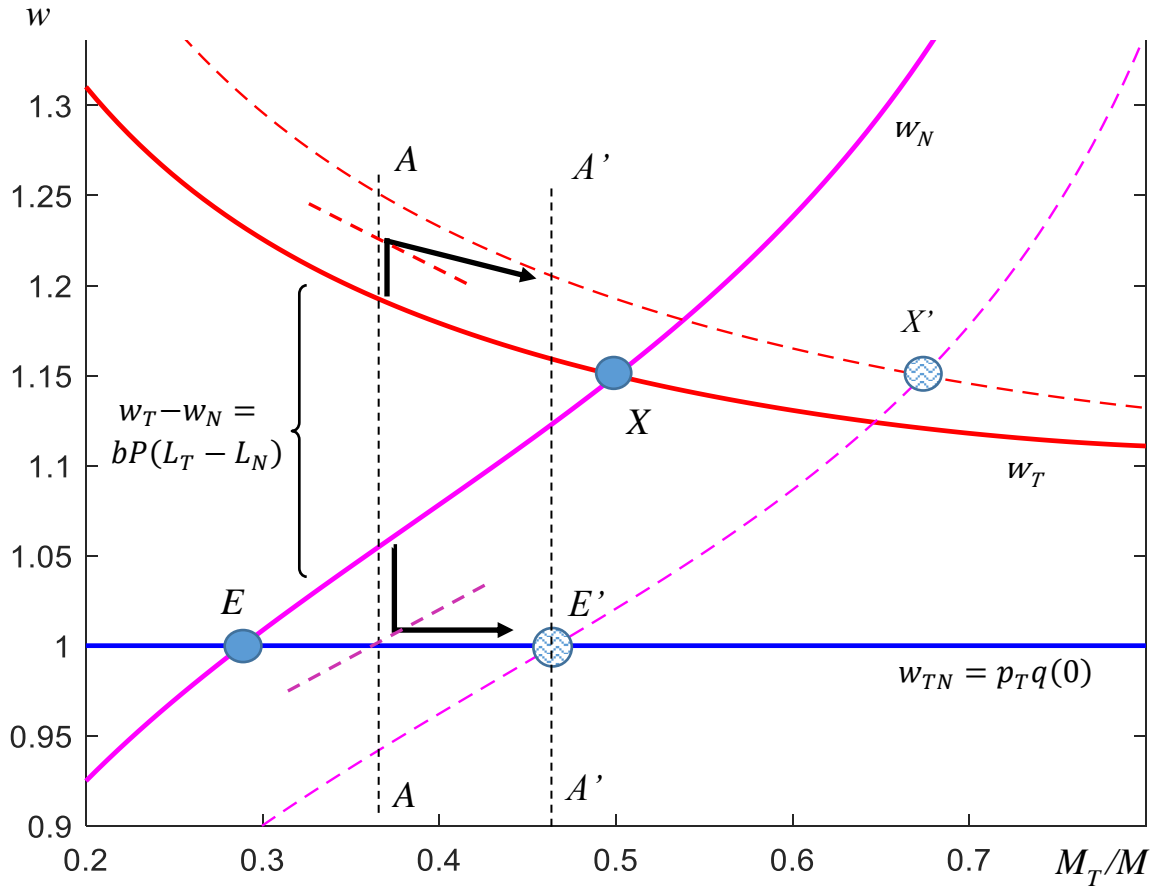
We now turn to fiscal policy, taking as example a reduction in net government spending which, in terms of the model, reduces G from zero to $G < 0$. The direct effect of this is to reduce demand for non-tradables, reducing their price and shifting the w_N schedule downwards. Given the simultaneous form of the system there is also a shift in the w_T curve, this moving upwards; the intuition is that, at given M_T/M , a lower value of w_N reduces L_N and therefore increases L_T which, with agglomeration economies, raises productivity and wages in type-T cities. Fig. 2 illustrates the combined effect, which amounts to a rightwards shift of both these wage curves to the dashed lines. The edges of the equilibrium set, points E and X , shift horizontally to the right to points E' and X' (see appendix).

If the economy is initially in the interior of the equilibrium set $[E, X]$ at AA , what happens as a fiscal contraction occurs? Adjustment can take two different forms depending on the magnitude of the contraction. In the first (which we call regime 1), a small $G < 0$ shifts w_N and w_T slightly to the right of the original solid lines, so AA remains between the (slightly right-shifted) intersections E and X . Since AA is not shifted, no cities switch activity. However, the intersections of the shifted w_T and w_N curves with AA move vertically up and down respectively (the vertical sections of the bold kinked-arrows on the AA line). The response to this widening wage gap between city types is migration out of type-N cities and into type-T; land rents, house prices, and urban costs in type-N cities fall, while type-T cities boom, with increasing population and land-rents.

⁸ See Venables (2018) for a fully specified model with two tradable sectors.

Larger reductions in G shift the wage curves and points E and X , further to the right. Regime 2 starts as E crosses line AA , i.e. the point where w_N has fallen to value w_{TN} , (at the kink in the lower bold arrow). At this point wages in type-N cities have fallen far enough to trigger tradable production even in places with no initial agglomeration benefits, so some type-N cities switch to becoming type-T. Line AA , giving the equilibrium value of M_T , is therefore dragged to the right and – at the final position of the dashed lines on Fig. 2 – the new division is at value E' , the intersection of w_{TN} with the new (dashed) w_N curve.

Figure 2: Contractionary fiscal policy

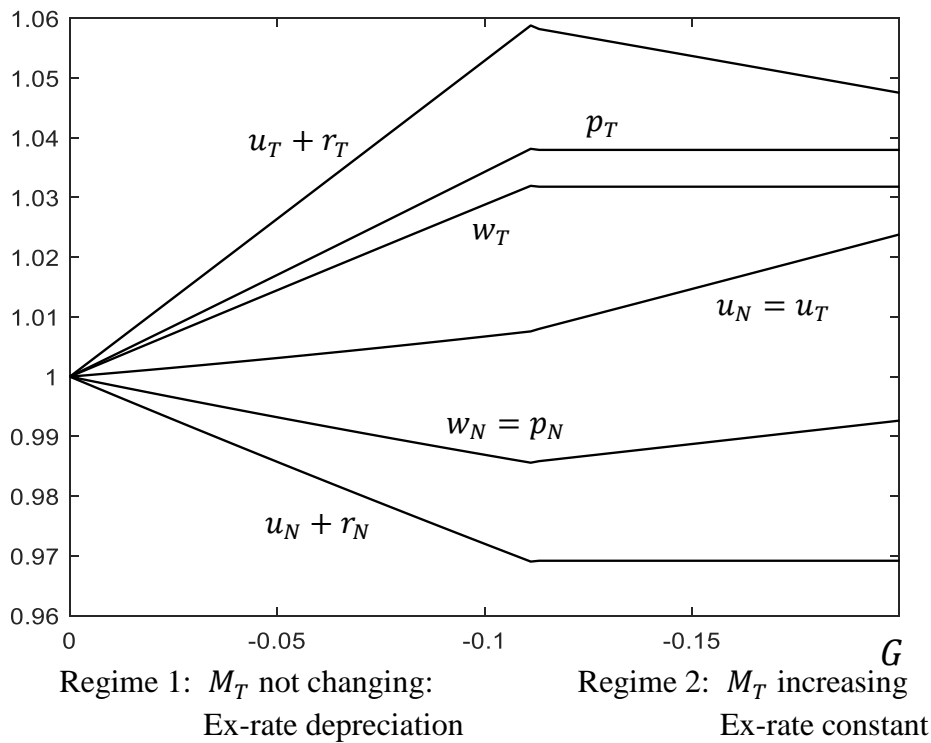


In both of these regimes fiscal contraction causes total employment in type-N cities to fall, but in regime 1 this is due to contraction of each type-N city, given the number of such cities; in regime 2 it involves some type-N cities switching to type-T. The difference between regimes is due to the fact that it takes a large fall in w_N to make the type-N places competitive for tradable goods production. Up to this point the effect of fiscal contraction is just to depress type-N cities, while existing type-T cities are experiencing increasing wages, population growth, and rising land and house prices,

What determines which of these adjustment mechanisms operates at particular levels of fiscal policy? In terms of the model, it depends partly on where you start; points E and X shift, and how far they have to move before triggering second stage adjustment depends on the initial position, AA . In practical terms, the question is, what does it take to trigger new tradable activities in ‘type-N’ cities that do not have the advantage of agglomeration or localisation economies? Evidently, this answer is sector and place specific, and depends on numerous social as well as economic factors. We return to it in following sections.

A summary of the behaviour of other variables in the model is given on Fig. 3. This shows values of variables relative to their value at $G = 0$, and is expressed in real rather than nominal terms, i.e. with price index P used as numeraire in the figure. One way of thinking about this is that the central bank is pursuing a policy of domestic price stability. Whereas in Figs. 1 and 2 the foreign price of tradables is held constant, now P is held constant and domestic currency prices p_T, w_N adjust to satisfy $P = p_T^{1-\theta} w_N^\theta = 1$; p_T can therefore be interpreted as the exchange rate. The variables recorded in Fig. 3 are the exchange rate, p_T ; the utility of each worker (the same regardless of location, $u_T = u_N$); the wages of a worker in each city, w_T, w_N ; and utility plus land-rent per capita in each city, $u_T + r_T, u_N + r_N$, this measuring the real income generated per worker in each city.⁹

Figure 3: Fiscal contraction: utility, real wages and rents per capita relative to values at $G = 0$.



⁹ Employment by city-type as a function of G is shown in appendix figure A1.

Fig. 3 makes clear the distinction between the two different adjustment mechanisms. In regime 1 fiscal contraction hits type-N cities, reducing w_N , causing out-migration from these cities and divergence of real income per worker $u_T + r_T$, $u_N + r_N$. As the exchange rate depreciates so type-T cities expand, but the depreciation is not large enough to trigger entry of tradables production in type-N cities. It takes a large fiscal contraction and exchange depreciation to trigger this entry, regime 2. Some type-N cities switch, raising their wages and productivity, and remaining type-N cities cease diverging from type-T. There is some convergence, although the net effect of fiscal contraction is to widen disparities between type-N and type-T cities.

6. Extensions

The model contains just one sort of heterogeneity – the distinction between tradable and non-tradable sectors – which economic behaviour maps into a heterogeneous urban structure. Clearly, in the world there is heterogeneity in numerous other dimensions, several of which we now discuss.

Heterogeneous labour. Tradable and non-tradable sectors will generally use different labour skills, so changes in prices and in the relative sizes of the two sectors will be associated with changes in the wages of different skill levels. The simplest setting is that there are two skill types (skilled and unskilled), each mobile between cities. They are employed in different proportions in the two sectors, and we suppose that the tradable sector is more skilled-labour intensive than the non-tradable.

Contractionary fiscal policy shrinks non-tradable production and expands tradables thereby raising demand for skilled labour relative to unskilled and increasing both the relative wage and the utility of skilled workers relative to unskilled. Thus, contractionary fiscal policy leads to divergence of wages between skill types, as well as between city types.¹⁰

Import competition (as described in section 3) has the opposite effect – type-T cities switch to type-N, this releasing skilled labour and raising demand for unskilled. This seems counter-intuitive, probably because of the assumption that there are just two skill types, so skilled labor released from a tradable sector that has been lost is perfectly equipped for skilled employment in remaining tradable sectors. An alternative model case where, for example, redundant skill types become effectively unskilled, could be easily modelled in a Roy-type assignment model.

Heterogeneous cities. We have assumed that cities are identical in fundamentals, with heterogeneity emerging from their specialisation. It follows that all type-T cities are identical, as are all type-N, and it is this that gives the two-regime adjustment; no type-N city switches to type-T until the point where all type-N cities are at this same switching point. If cities varied

¹⁰ Mobility ensures that the utility of a worker of each type is the same in all cities. See the appendix for more detail on the two-skill model.

in their attractiveness to tradable sector firms, then the least unattractive would switch at a smaller wage differential ($w_T - w_N$) than the more unattractive. Adjustment would still be a combination of type-N cities contracting, and type-N cities switching to type-T, but this would not necessarily be in two distinct regimes. In terms of Fig. 3, the single kink would be removed. If adjustment were to alternate between the two mechanisms, then the lines would alternate their first and second regime gradients.

Urban housing supply and migration. Large cities are expensive, captured in the model by the relationship between city size and the urban costs of commuting and rent, bPL . The scale of population movement in response to wage differences is smaller the larger is parameter b , this corresponding to low supply elasticity of housing. More generally, urban costs depend on the shape of the city, building technologies, and the price elasticity of household demand for space.¹¹ Migration depends on individuals' heterogeneous moving costs and preferences over place of residence. These may dampen migration responses, while creating a problem of those left-behind.

Further feedback mechanisms. The paper models the consequences of a single feedback mechanism, the positive one associated with agglomeration in tradable sectors. In reality further feedback mechanisms operate, many of them negative feedbacks in cities that have lost jobs in traditional sectors. They include loss of fiscal revenue and associated decline in the quality of public services; selective migration changing the age and skill composition of the population; deterioration of educational opportunities and aspirations; and, in some cases, loss of identity and self-worth. They all reinforce the premise of this paper, that it may be difficult to get new, internationally competitive, activities started in a place that has experienced a negative shock.

7. Concluding comments

We have argued four main points. First, the core of the economic problem is that the market mechanism does not create sufficient incentives to start new tradable activities (or more generally, new activities that can achieve high productivity through returns to scale and agglomeration economies) in places that have lost historic specialisms. This is not because of price rigidities or policy distortions. It is simply because of technologies that create agglomeration economies and the consequent 'first-mover' problem.¹² These effects are likely

¹¹ If the city is a circular disk, all households occupy the same amount of land, and commuting costs are proportional to distance, then costs increase with the square root of L . Some general statements are given in Henderson and Venables (2009).

¹² This is in sharp contrast to much international economics that puts naive faith in adjustment occurring because 'everywhere has a *comparative* advantage'. This statement is true, but operative only if firms' costs can vary sufficiently across space for comparative advantage to become competitive advantage. Within a country prices of immobile factors (land and housing) can diverge but – if they cannot go negative – do not have sufficient leverage to enable comparative advantage to operate.

to be augmented by further considerations that deter new activities from moving into areas that are perceived to be declining. Such areas face out-migration of skills, loss of tax revenue and public services, as well as demographic and other social disadvantages all of which make them unattractive to inwards investment.

Second, the adjustment mechanism in response to trade shocks is more likely to take the form of cities that lose tradable activity switching to non-tradables, than it is of them acquiring new tradable sectors. This comes over in a very stark way in the model because of the assumption that tradability and agglomeration go together. Of course, this is not exact; some tradable goods do not cluster, and some non-tradables do. But this assumption seems to hold good for many of the sectors that we see in booming cities.

Third, import competition increases polarisation of the urban system. The relative price of non-tradables falls, and with it the wages of people in towns and cities producing such goods. Mobile factors will move in response to this loss, so the negative impact is transferred to factors of production that cannot move. In this simple model this is just land, but in reality includes individuals who are unable or unwilling to move. The social, economic, and political consequences of this are apparent.

Fourth, fiscal policy will have different impacts on cities producing tradable, vs non-tradable goods. Contractionary policy has two effects; one is to shrink the size of existing non-tradable cities; the other is to shrink the number of non-tradable cities. They have very different implications, the former causing divergence of city-types, the latter enabling some of these cities to switch to higher productivity tradable sector jobs. Given agglomeration economies and the attendant first-mover problem, the former mechanism is the one more likely to operate.

What can facilitate the start-up of new tradable and agglomeration prone activities? One possibility is based on innovation. If technology creates new activities, not linked to existing agglomerations, then it is possible that they start up in relatively low cost ‘type-N’ places. This possibility is emphasised in Moretti (2013), and exemplified by Seattle. In the 1970s Seattle was a city with a declining port and manufacturing sector, unemployment twice the US national average, losing population and famous for the 1971 billboard saying ‘will the last person leaving Seattle turn out the lights’. Microsoft arrived, losing little if any productivity in moving from its original base of Albuquerque. The cluster of software activity then grew up around Microsoft. Of course, the move of Microsoft from Albuquerque was due to the fortuitous circumstance that both Bill Gates and Paul Allen had grown up in Seattle. While the innovation route has been successful in transforming some cities, it seems unlikely that there are enough distinct new innovative clusters for this to be a solution for more than a few fortunate cities.

A second route is to try to address the coordination failure by policy that targets particular places for economic development, possibly in specific sectors. Economic reasoning suggests that ‘large developers’ may be able to internalise the externalities created by agglomeration, overcoming coordination failure by launching development at scale. Public policy to support this may take the form of city plans and the location of infrastructure (e.g. placement of transport hubs). Special economic zones offer regulatory, fiscal, and infrastructure benefits, concentrated in one place with the hope of creating cluster benefits. Developing countries offer some successful example, such as Shenzhen, Dhaka, and Penang.¹³ However, there are many more failures. Developed countries have used fiscal incentives in the form of regional investment or employment subsidies and subsidies to influence plant location decisions. Reviews of such policies suggest that, even if policies have had some impact, they have generally failed to jump-start new economic activities and trigger the development of self-sustaining private sector clusters (see e.g. Neumark and Simpson 2015, Kline and Moretti 2014). Perhaps one reason for this is that such policies succeed in attracting non-tradable activities, moving public sector jobs or securing investments in warehousing or customer service centres; policies may even be targeted at these sectors. Such policies fail to attract internationally competitive tradable sectors, so their effect is to cause displacement of non-tradable activity and accelerate the adjustment process described in this paper, dragging down incomes in all type-N cities.

A third possibility is that parts of tradable sector activities are able to split off from their core function in type-T cities and relocate to a lower wage type-N city. This brings the dual benefit of job creation in type-N cities and creating space for core activities to expand in type-T cities. The extent to which this is possible depends on the scope – functional and spatial – of agglomeration economies. Internationally, the growth of offshoring indicates that it is possible to geographically separate back-office activities from parts of the business that benefit from presence in a cluster (e.g. in some finance and insurance sectors). However, two caveats are in order. First, the international context is one of much larger nominal wage differences than those that arise between cities within a country. And second, the move to ‘reshoring’ is illustrative of the fact that many firms found the costs of geographical fragmentation of activities to be greater than anticipated.

The final option is to accept increasing polarization of the urban structure and the attendant decline of many towns and cities. In a simple framework – such as that presented in this paper – this could be welfare improving, particularly if booming cities are enabled to expand by constructing infrastructure and housing to mitigate effects on commuting costs, land prices and rents.¹⁴ However, the further costs are evident. Many people are unable or unwilling to move,

¹³ See Duranton and Venables (2020) for analysis of place-based policies developing economies.

¹⁴ Hsieh and Moretti (2018) argue that restrictions on housing construction in booming urban areas has reduced aggregate US growth by 1/3rd over the period 1964-2009.

with economic, social and health costs that are apparent. Mitigating these costs is important, and we note that there is a significant resource that could be used to finance this. As we saw above, the beneficiaries of urban polarisation are those fortunate to receive windfall gains from owning land in the booming cities, gains which are administratively, economically, and ethically – if not politically – ripe for taxation.

References:

- Combes, P-P. and L. Gobillon (2015). The Empirics of Agglomeration Economies. In G. Duranton, J.V. Henderson, and W. Strange (eds.) *Handbook of Regional and Urban Economics*, volume 5A, 247–348. Elsevier, Amsterdam.
- Duranton, G. and D. Puga, (2004). ‘Micro-foundations of urban agglomeration economies’, in: J. V. Henderson and J. F. Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4, 2063-2117. Elsevier, Amsterdam.
- Duranton, G. and D. Puga, (2015). ‘Urban Land Use’, in: G. Duranton, J. V. Henderson and W. Strange (eds.), *Handbook of Regional and Urban Economics*, volume 5, chapter 8, pages 467-560, Elsevier, Amsterdam.
- Duranton, G. and A.J. Venables (2020), ‘Place-based policies: principles and developing country applications’, *Handbook of Regional Science*, eds. M. Fischer and P. Nijkamp, Springer.
- Glaeser E. L. (2012) *The Triumph of the City*, Macmillan, London
- Gyourko, J., Mayer, C., Sinai, T. (2013) ‘Superstar Cities’, *American Economic Journal: Economic Policy*, 5(4), 167–99.
- Haldane, A. (2019) ‘Is all economics local?’ Bank of England, London,
- Henderson, J. V. (1974). ‘The Sizes and Types of Cities’, *American Economic Review* 64(4), 640–656.
- Henderson, J.V. and A.J. Venables (2009), ‘The dynamics of city formation’, *Review of Economic Dynamics*, 12, 233-254.
- Hsieh, C.-T. and E. Moretti (2019) ‘Housing constraints and spatial misallocation’, *American Economic Journal; Macroeconomics*, 11.2, 1-39.
- Kline, O. and E. Moretti, (2014) People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Policies *Annual Review of Economics*, 6, 629-662.
- Moretti, E. (2013) *The New Geography of Jobs*, Mariner Books, New York.
- Neumark, D. and H. Simpson. (2015). ‘Place-based policies’ In: G. Duranton, J. V. Henderson, and W. C. Strange, *Handbook of Regional and Urban Economics*, volume 5B. Amsterdam: Elsevier, 1197-1287.
- Rosenthal S. & W. Strange. (2004). ‘Evidence on the nature and sources of agglomeration economies’, in V. Henderson & J-F. Thisse (eds) *Handbook of Regional and Urban Economics*, vol 4, pp 2119-2171. Elsevier, Amsterdam.
- Storper, M. (2018) ‘Separate Worlds? Explaining the current wave of regional economic polarization’, *Journal of Economic Geography*, 18, 247-270
- Venables, A. J. (2018) ‘Globalisation and Urban Polarisation’, *Review of International Economics*, 26, 981-996.

Appendix:

Equilibrium: Set $p_T = 1$, so eqn (4) is $P = w_N^\theta$. Equations (2) and (5) can be solved for values of L_T, L_N , so $L_T = [(M - M_T)(w_T - w_N) + PbL]/PbM$, $L_N = [PbL - M_T(w_T - w_N)]/PbM$. Eqns (1) and (3) are $w_T = p_T q(L_T)$, $w_N = \theta[M_T L_T w_T + G]/[(1 - \theta)(M - M_T)L_N]$. Substituting values of L_T, L_N , these are two non-linear simultaneous equations in parameters of the model, w_T, w_N , and M_T .

Figures in the paper trace solutions numerically, for different values of M_T , and for parameter values $L = 100, M = 100, \theta = 0.5, b = 0.3, q(L_T) = 1 + 0.015L_T$. The intersection point X is at $M_T/M=0.5$ because $\theta = 0.5$.

Rent: Total real rent is $R = M_T bL_T^2 + (M - M_T)bL_N^2$. If this were linear in L_T, L_N it would be constant (since full employment, eqn. 2 is $L = M_T L_T + (M - M_T)L_N$). The quadratic terms induce strict convexity with a minimum point at X, where cities are all the same size.

Fiscal policy: At point X, $w_T = w_N$ and $L_T = L_N$. This means that the total differential of eqns. (1) – (5) with respect to G gives $dM_T/dG = -\theta/((1 - \theta)w_N L_N)$, and all other variables unchanged. In particular, $dw_T/dG = dw_N/dG = 0$ indicating that wages are unchanged as point X moves, i.e. X is shifted horizontally on Fig 2.

Heterogeneous labour: There are fixed endowments of the two skill types, and production functions in each sector using labour of both types. Migration equations hold for workers of both types, and we assume that urban costs are a function of the total population of the city. Eqn. (5) therefore becomes, $w_{TS} - bP(L_{TS} + L_{TU}) = w_{NS} - bP(L_{NS} + L_{NU})$, $w_{TU} - bP(L_{TS} + L_{TU}) = w_{NU} - bP(L_{TS} + L_{TU})$.

Figure A1: Fiscal contraction: employment by city type, relative to values at $G = 0$.

