DISCUSSION PAPER SERIES

DP11212

BREAKING INTO TRADABLES: URBAN FORM AND URBAN FUNCTION IN A DEVELOPING CITY

Anthony J Venables

DEVELOPMENT ECONOMICS

ISSN 0265-8003

BREAKING INTO TRADABLES: URBAN FORM AND URBAN FUNCTION IN A DEVELOPING CITY

Anthony J Venables

Discussion Paper 11212
Published 04 April 2016
Submitted 04 April 2016

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

BREAKING INTO TRADABLES: URBAN FORM AND URBAN FUNCTION IN A DEVELOPING CITY

Abstract

Many cities in developing economies, particularly in Africa, are experiencing ‘urbanisation without industrialisation’. This paper conceptualizes this in a framework in which a city can produce non-tradable goods and – if it is sufficiently competitive – also internationally tradable goods, potentially subject to increasing returns to scale. The city may get locked into non-tradable production for three distinct reasons. One is high demand for non-tradables, arising if the city receives income from other sources (such as natural resource rents), creating an urban Dutch disease. A second is coordination failure amongst potential producers of tradable goods. A third arises from the way in which the city is built. In a two-period model with sunk construction costs, expectations about second period outcomes shape construction decisions, which in turn shape the competitiveness of the city. There may be two (perfect foresight) equilibria. One in which land values are relatively low, little is invested in buildings, and the city is too low density and too high cost to attract tradable production. The other in which land values are high, the city is built taller and denser and is able to accommodate more workers and attract tradable production; this, in turn, generates the employment and income to support high land values. The former configuration is consistent with the ‘urbanisation without industrialisation’ experience of many African cities.

JEL Classification: O14, O18, R1, R3

Keywords: city, urban, economic development, tradable goods, structural transformation.

Anthony J Venables - tony.venables@economics.ox.ac.uk
University of Oxford and CEPR

Acknowledgements
We gratefully acknowledge the support of an Africa Research Program on Spatial Development of Cities at LSE and Oxford funded by the Multi Donor Trust Fund on Sustainable Urbanization of the World Bank and supported by the UK Department for International Development. Thanks also to Vernon Henderson, Ferdinand Rauch, Paul Collier, Somik Lall, Patricia Jones and seminar participants for helpful comments. Opinions expressed in this paper are those of the author and do not represent those of the World Bank

`

23/3/2016

Breaking into Tradables:

urban form and urban function in a developing city

Anthony J. Venables
University of Oxford

Abstract

Many cities in developing economies, particularly in Africa, are experiencing 'urbanisation without industrialisation' This paper conceptualises this in a framework in which a city can produce non-tradable goods and – if it is sufficiently competitive – also internationally tradable goods, potentially subject to increasing returns to scale. The city may get locked into non-tradable production for three distinct reasons. One is high demand for non-tradables, arising if the city receives income from other sources (such as natural resource rents), creating an urban Dutch disease. A second is coordination failure amongst potential producers of tradable goods. A third arises from the way in which the city is built. In a two-period model with sunk construction costs, expectations about second period outcomes shape construction decisions, which in turn shape the competitiveness of the city. There may be two (perfect foresight) equilibria. One in which land values are relatively low, little is invested in buildings, and the city is too low density and too high cost to attract tradable production. The other in which land values are high, the city is built taller and denser and is able to accommodate more workers and attract tradable production; this, in turn, generates the employment and income to support high land values. The former configuration is consistent with the 'urbanisation without industrialisation' experience of many African cities.

Keywords: city, urban, economic development, tradable goods, structural transformation.
JEL classification: O14, O18, R1, R3

Acknowledgements: We gratefully acknowledge the support of an Africa Research Program on Spatial Development of Cities at LSE and Oxford funded by the Multi Donor Trust Fund on Sustainable Urbanization of the World Bank and supported by the UK Department for International Development. Thanks also to Vernon Henderson, Ferdinand Rauch, Paul Collier, Somik Lall, Patricia Jones and seminar participants for helpful comments. Opinions expressed in this paper are those of the author and do not represent those of the World Bank.

Author's Address:
A.J. Venables
Department of Economics
Manor Road
Oxford OX1 3UQ UK
tony.venables@economics.ox.ac.uk

1. Introduction

Cities in the developing world face the challenge of accommodating a predicted 2.5bn more people by 2050, and of creating jobs for this population. The fastest urbanising region is Sub-Saharan Africa (SSA) where the urban population is predicted to treble to more than 1bn by 2050. Historically, urbanisation has been accompanied by substantial job creation in tradable goods sectors, including manufacturing, and some countries, notably in East Asia, have followed this path in recent years. However, cities in many other countries have failed to attract tradable-sector jobs. The phrase 'urbanisation with industrialisation' has gained currency, and Gollin et al. (2015) point to the prevalence of this phenomenon in resource rich economies. There is evidence that this is a widespread feature of African urbanisation. In the cross-section of non-African economies the manufacturing share of GDP rises from 10% to nearly 20% as the urban population share rises to 60%, beyond which it falls back. In Africa the manufacturing share remains flat (or somewhat falling) at around 10% of GDP as urbanisation reaches these levels (Jones 2016). World Bank enterprise surveys report the share of firms producing tradable vs non-tradable goods in cities around the world. The share producing tradables in a (small) sample of Asian and Latin American cities is around 85%. In a sample of African cities it averages just 60%.[1]

The objective of this paper is to develop a framework in which to conceptualise 'urbanisation without tradable sectors', and thence draw out implications for city performance and policy. While there are many reasons why African countries are failing to attract internationally footloose activity, this paper focuses on one particular aspect – the characteristics of African cities.

The link we seek to make is between urban form and urban function. By urban form, we mean the physical, spatial and organisational characteristics of the city, in so far as these factors shape the incremental costs associated with urban size. The model we develop is based on the standard model of the open city, able to attract labour from outside at constant real wage. While the real wage is constant, the nominal urban wage must be higher in larger cities in order to compensate for higher urban costs. These costs are endogenous, and are lower if cities have efficient land use and infrastructure, this supporting a dense and well-connected urban form.

By urban function, we mean what cities do, in particular whether they are able to attract investment in tradable sectors or whether this avenue is closed off, forcing specialisation in non-tradable activities. To capture this we augment the standard urban model by adding the possibility that the city can perform different productive activities. One is production of non-tradables, for consumption in the city or perhaps more widely in its domestic hinterland. The

[1] Jones (2016). Firms with more than 5 employees.

other is tradable production for wider world markets. The critical difference between sectors is in the returns to expanding employment. Non-tradables have diminishing returns, perhaps for technological reasons, but more fundamentally because the price of non-traded output is set locally and diminishes as supply increases. In contrast, tradables face a perfectly elastic world demand curve and may also be subject to agglomeration economies, in which case there are increasing returns to employment in the sector. [2]

Combining these ingredients generates several possible equilibrium configurations, one of which appears to be common in African cities. This is that real wages are low, but nominal wages (i.e. wages at market exchange rates) are relatively high, making cities uncompetitive in tradable goods and forcing specialisation in non-tradables. There are two sides to this configuration. One is a high cost of living, due in part to urban costs (and hence 'urban form'); the nominal wage must therefore be relatively high to attract a supply of labour. The other is a relatively high price of non-tradable output, set by local supply and demand; this enables firms in the non-tradable sector to pay high nominal wages. While labour productivity is low in both sectors, this can be passed on as higher prices in the non-tradable sector, but not in the tradable sector which is therefore inactive.

This configuration is confirmed by several studies.[3] Firms in African cities pay wages (at official exchange rates) about 15% higher than in non-African cities, conditional on national GDP pc. Labour costs are estimated at up to 50% higher. Corresponding to this, sales per worker are about 25% higher than in comparable non-African cities, but this is largely in the non-traded sectors and appears to simply reflect higher prices. The higher cost of living shows up in high rents, as well as in prices of food and other goods.

Our core model is outlined in the next section of the paper, and we show how non-tradable production crowds out tradable activity. Following sections go on to show that the combination of diminishing returns in non-tradables and increasing returns in tradables creates the possibility of multiple equilibria. There is a low-level trap with the configuration described above, and also an equilibrium in which tradable production operates at sufficient scale to achieve agglomeration economies, raising productivity and city size. These outcomes occur at different parameter values, but there is also a range of values in which both are equilibrium outcomes. Thus, cities with identical fundamentals can have quite different outcomes. The source of the low-level trap is a standard coordination failure, in which it is not worthwhile for any firm to commence tradable production, absent the

[2] We think of tradables as products such as textiles, apparel, or participation in global value chains, whereas non-tradables are beer, cement, retailing. This is different from the notion of non-tradable business services or intermediates in, for example, Duranton and Puga (2005).

[3] This paragraph draws on Jones (2016).

productivity benefits derived from operating the sector at scale.[4] The consequences are smaller city size and potentially large welfare losses.

Final sections of the paper argue that the coordination failure has deeper foundations in the way in which the city gets constructed. The physical structure of the city is built by sinking capital in long-lived buildings, and therefore depends on expectations of the future performance of the city. In a two-period extension of the model we show that multiple equilibria – each with perfect foresight – can occur in the way in which the city is constructed and its consequent pattern of specialisation. The low-level equilibrium has a city with low land rents, low density housing, low population, and produces only non-tradables in both periods. By contrast, in the high-level equilibrium second period urban rents and house prices are expected to be high. As a consequence first period building – on a larger scale and at a higher density – is profitable. This expands the city, shifting the labour supply curve to the right, so giving lower nominal wages (for each level of employment) and triggering the start of tradable goods production. Expectations are therefore self-fulfilling.

The model is constructed to focus on the interplay between tradable and non-tradable sectors, urban costs, and expectations. Developing country cities also contain many market and policy failures. While these are not explicitly captured in our model, the concluding section of the paper outlines ways that they could be represented in the model and their likely impact on results. It suggests that they reinforce the conclusions of the formal model.

2. The model

The model is developed using two relationships and, where possible, a diagrammatic approach. One relationship is the city's demand for labour, i.e. the relationship between the wage and the level of employment in the city, this depending on 'urban function'. The other is the supply of labour to the city, i.e. the relationship between the wage and city population, this given by migration and depending on the cost of living in the city and hence 'urban form'.

Production and labour demand: The population of the city is endogenous and denoted L. Labour is the only factor of production and can be used in either a tradable or a non-tradable activity, with employment levels satisfying $L_T + L_N = L$. Labour productivity in the tradable sector is $a(L_T)$, and we assume that this is either constant or increasing in L_T, increasing returns arising because of agglomeration economies external to the firm but internal to the tradable sector. The price of the tradable good is fixed on world markets and we use it as

[4] See Murphy et al. (1989), Henderson and Venables (2009).

numeraire, so the wage offered in the sector is $a(L_T)$. The sector operates if this is greater than or equal to the market wage, *w*, so the following relationship must hold,

$$w = a(L_T),\ L_T \geq 0; \qquad w > a(L_T),\ L_T = 0. \tag{1}$$

The non-traded sector is always active, has price p_N, and uses labour alone under conditions of constant returns to scale, giving (with units chosen such that one unit of labour produces one unit of output)

$$x_N = L_N = L - L_T, \qquad w = p_N. \tag{2}$$

The price of non-tradables equates supply and demand, this taking the form $(L - L_T)p_N = \mu wL + p_N h(p_N)$. The left-hand side is the value of output, and the right the value of demand for non-tradables, containing two terms. The first is demand derived from income generated in the city; city income is the wage bill, fraction μ of which is spent on non-tradables. The second term, $p_N h(p_N)$, is spending from income generated outside the city. Thus, the function $h(p_N)$ is demand for non-tradables derived from resource rents and tax revenues that are transferred to the city, plus hinterland demand for non-tradables produced in the city. We call this the hinterland demand function, and take it as exogenous and decreasing in price. Using $w = p_N$, non-tradable market clearing can be rearranged as

$$w = h^{-1}([1-\mu]L - L_T). \tag{3}$$

This is the price (= wage) at which net urban supply equals hinterland demand, where $h^{-1}(.)$ is the inverse hinterland demand curve. Together, relationships (1) and (3) implicitly define the city's (inverse) labour demand schedule, i.e. they give the value of *w* at which urban employment (in non-tradables and tradables) fully employs a labour force of size *L*. The form of this demand schedule is central to much of what follows, and it can be summarised as follows.

If the tradable sector is non-active, then $L_T = 0$ and, from (3),

$$w^D(L : L_T = 0) = h^{-1}([1-\mu]L). \tag{4a}$$

If the tradable sector is active, then $L_T > 0$ and, from (1) and (3),

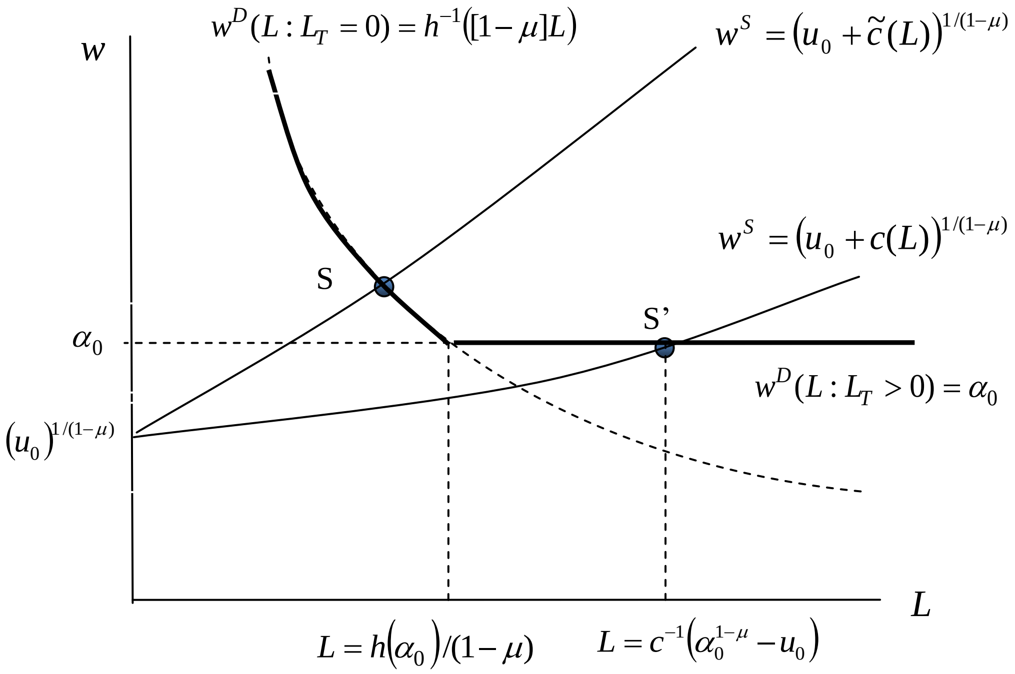
$$w^D(L : L_T > 0) = a(L_T), \text{ with } L_T \text{ solving } a(L_T) = h^{-1}([1-\mu]L - L_T). \tag{4b}$$

The wage offered is the maximum,

$$w^D = \max\left[w^D(L : L_T = 0), w^D(L : L_T > 0)\right]. \tag{4c}$$

This is illustrated by the bold curve in Figure 1 which has city population on the horizontal axis and the wage rate on the vertical. The figure is drawn for the simplest case in which there are constant returns to scale in tradable production so labour productivity in tradables is a constant α_0, $a(L_T) = \alpha_0$. On the downward sloping segment the entire labour force is employed in the non-tradable sector, $w^D = h^{-1}([1-\mu]L)$, downwards sloping as more employment increases supply of the non-traded good, reducing price and wages. If the ensuing wage is less than α_0 then tradable production is profitable, giving the horizontal segment, $w^D = \alpha_0$. The dividing line is where the urban population is $L(1-\mu) = h(\alpha_0)$.

Figure 1: Urban specialisation with constant returns to scale



Urban population and labour supply: Urban households spend income on a Cobb-Douglas composite of non-tradable and tradable goods, with respective shares μ and 1- μ. This composite good crops up in much of what follows, and its price index is w^{μ} (since tradables are the numeraire and the price of non-tradables is the wage). Of this spending $c(L)w^{\mu}$ goes on urban costs, i.e. purchasing a package of urban necessities (transport, housing) the quantity of which is $c(L)$ and the price of which is w^{μ}, since we assume that the urban necessities themselves use the composite good. $c(L)$ is assumed to be increasing in city population, and in section 4 we derive this from the spatial structure of the city. These costs are necessary for urban living, and utility is derived from income net of these costs. An urban household's utility is therefore $\left[w - c(L)w^{\mu}\right]w^{-\mu}$. The term in square brackets is the value of consumption net of urban costs, and this is deflated by w^{μ}, the cost of the composite good.

City population, L, is determined by the possibility of migration from the rest of the country, where per capita utility is constant at u_0. Migration occurs until urban utility equals u_0, i.e.

$$u_0 = \left[w - c(L)w^{\mu}\right]w^{-\mu} = w^{1-\mu} - c(L). \tag{5}$$

Rearranging, the (inverse) labour supply schedule facing the city, i.e. the urban wage w^S required to attract population L is

$$w^S = \left(u_0 + c(L)\right)^{1/(1-\mu)}. \tag{6}$$

This is illustrated by the upwards sloping curve on figure 1, drawn for two different urban cost functions.

Urban Dutch disease: equilibrium with constant returns in tradable production: Urban labour supply and demand, equations (4) and (6), give equilibrium values of the wage, city size and sectoral employment, as illustrated by points S and S' on figure 1. Point S gives the equilibrium if urban costs are high, $\tilde{c}(L)$, meaning that the city produces only non-tradables. At lower costs equilibrium is at S', with both sectors active and wage rate $w = \alpha_0$. City population and employment in tradables come from (6) and (4b), respectively, $L = c^{-1}\left(\alpha_0^{1-\mu} - u_0\right)$, $L_T = (1-\mu)L - h(\alpha_0)$.

The message is simple. A city may have low real wages u_0, but high urban costs mean that nominal wages are too high to attract footloose manufacturing industry, the tradable sector, since $w > \alpha_0 > u_0$, as at point S. This is more likely the higher are urban costs and the larger the demand for non-traded goods, as given by hinterland demand for non-tradables and μ.[5] This is an urban 'Dutch disease' phenomenon, as spending on non-tradables crowds out tradable production.

Given this outline of the model, in following sections we: (i) enrich the labour demand curve by adding increasing returns in tradable goods production; (ii) go behind the labour supply

---

[5] Notice that the city can exist without tradable production only if there is hinterland demand for its non-tradables or if it consumes no tradables, $\mu = 1$. Higher values of hinterland demand and of μ shift the kink in the labour demand schedule to the right. As $L \to \infty$ the share of tradables in employment rises to limiting value $1 - \mu$.

curve, adding commuting and construction costs and land rents and investigating real income and welfare; (iii) set out a two period model to address the importance of sunk construction costs, forward looking expectations, and urban density.

3. Increasing returns in tradable production

We now suppose that tradable production is subject to agglomeration economies, so that productivity in tradable production is increasing with the size of the sector.[6] For simplicity we make this relationship linear in tradable employment between lower and upper bounds, α_0, α_m, so

$$a(L_T) = \min[\alpha_0 + \alpha L_T, \alpha_m], \quad \alpha > 0, \quad \alpha_0 < \alpha_m. \tag{7}$$

The labour demand schedule is now as illustrated in figure 2. At $L \leq h(\alpha_0)/(1-\mu)$ all labour is employed in non-tradables, giving the downwards sloping segment as before (equation 4a). Beyond this point tradable production takes place, and labour is allocated between sectors to equate average products, i.e. from (4b) and (7), L_T solves

$$h^{-1}([1-\mu]L - L_T) = \min[\alpha_0 + \alpha L_T, \alpha_m]. \tag{8}$$

As tradable employment increases from zero so productivity rises, giving the upward sloping segment on figure 2.[7] When productivity reaches α_m the schedule becomes horizontal, and each further unit increase in L expands L_T by 1 - μ (from (8), with the right-hand side constant at α_m). Labour supply, w^S, is as before.

Figure 2 illustrates equilibrium configurations for three different levels of urban costs. If costs are high the city is specialised in non-tradables, as at S. If costs are low enough there is a unique equilibrium with both tradable and non-tradable production (S'). The wage is set by productivity in the traded sector, city size follows from labour supply, and employment in each sector comes from equation (4b).

There is an intermediate range of urban costs for which there are multiple equilibria, with the two stable equilibria labelled M and M'.[8] At the lower equilibrium, M, wages are too high

---

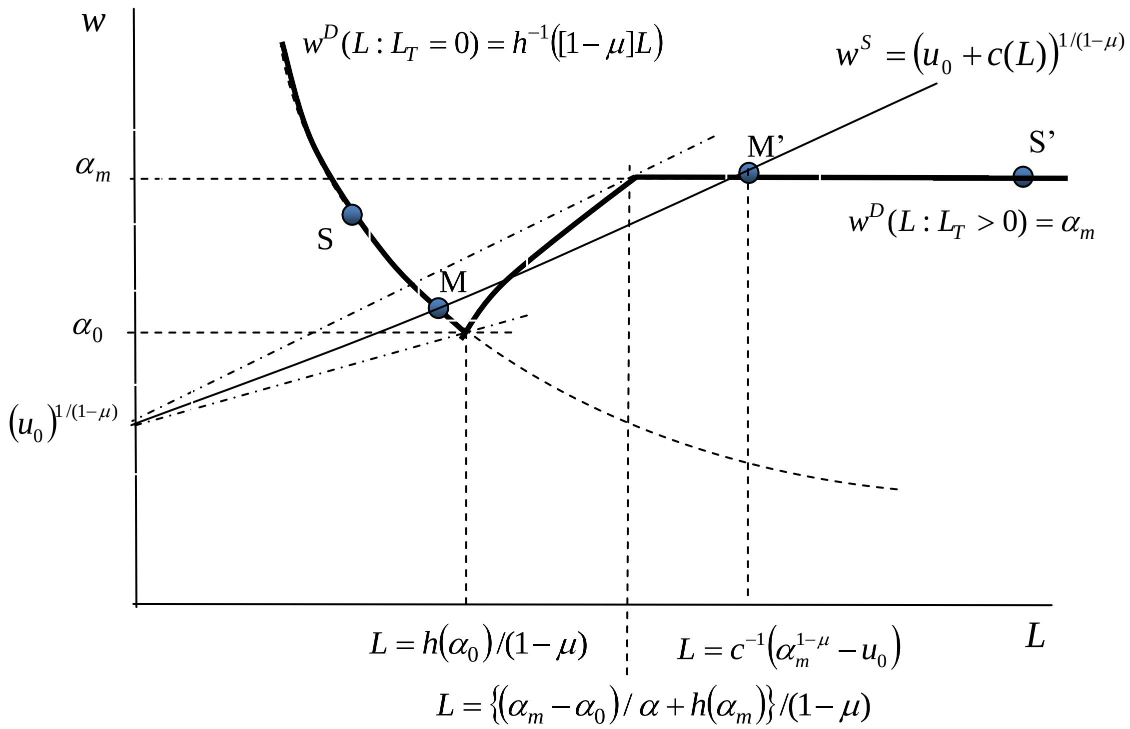
[6] For discussion of agglomeration economies in the context of developing economies see World Bank (2009).

[7] The gradient is, from (8), $dw/dL = \alpha dL_T/dL = \alpha(1-\mu)/(1+\alpha h')$, less than α as increasing city employment, L, also changes demand for non-tradables.

[8] There is a further equilibrium at the intersection between M and M' but this is unstable if tradable producers enter/ exit according to their profitability.

for entry of a small mass of tradable producers to be profitable ($w > \alpha_0$). At the upper one, M', the tradable sector is large enough for agglomeration economies to have cut in, raising productivity. The multiplicity of equilibria come from the fact that agglomeration economies are externalities, and no single producer is large enough to internalise the benefits in its entry decision. Coordinated entry of a large mass of producers would remove the low-level equilibrium at M but, given the coordination failure, it is not profitable for any single producer to enter at this point. Multiple equilibria arise only if the upward sloping section of the labour demand curve is steep enough, i.e. increasing returns are stronger than the diseconomies of scale created by urban costs. Conditions for this are set out in the appendix, as is a summary characterisation of the equilibrium points illustrated in the figure.

Figure 2: Increasing returns and multiple equilibria.



Several messages come from this. First, the urban economy has a region of diminishing returns to scale before it reaches the possibility of increasing returns. Given demand for non-tradables, it can reach the increasing returns segment only if urban costs are sufficiently low.

Second, coordination failure associated with agglomeration externalities may create multiple equilibria. Thus, a city may be locked into a low-level equilibrium even though – given its fundamentals – there is another equilibrium with higher employment. While the model – so far – is completely static, it is worth asking how to trigger a move from M to M'. Lower urban costs move M down and to the right, but it is not until the labour supply curve lies below the minimum point on the labour demand curve that uncoordinated entry of tradable producers becomes profitable. We pursue this further in a two period model in sections 5 and 6, but first add a richer spatial structure to the model.

4. Spatial structure and welfare:

To this point we have simply assumed that urban costs are an increasing function of urban population, $c(L)$. We now ground this in a model of urban space and residential construction in order to establish the foundations of the urban cost relationship and also to derive a welfare indicator. The structure will enable us to look (in section 5) at the implications of sunk construction costs and the role of expectations in shaping urban form.

Urban costs are derived from a spatial structure, following Alonso (1964) and others. We suppose that the city is linear (or comprised of a number of linear spokes). Workers are employed in the central business district (CBD) and incur commuting costs. These take the form of t units of the composite good per unit distance, so a worker living at distance x from the CBD pays commuting costs xtw^{μ}. Each worker occupies a 'house' that provides one unit of floor-space and, since workers are perfectly mobile, house prices at each distance, $p(x)$, must satisfy the indifference condition, $u_0 = \left(w - p(x) - xtw^{\mu}\right)w^{-\mu}$ (see equation 5), i.e.

$$p(x) = w - \left(u_0 + xt\right)w^{\mu} \ . \tag{9}$$

House prices consist of two elements, construction costs, and land rent. A unit of land can be built to density D at cost bw^{μ} (i.e. it takes b units of composite good to build D houses per unit land, D and b assumed exogenous until section 6). Rent per unit land, $R(x)$, is revenue from houses minus construction costs,

$$R(x) = p(x)D - bw^{\mu} . \tag{10}$$

The edge of the city, denoted $\hat{x}$, is where rent equals zero, $R(\hat{x}) = 0$. From (10), this is where the price of a house equals its construction cost, $p(\hat{x}) = bw^{\mu} / D$, and using this in (9), $\hat{x}$ is implicitly defined by

$$u_0 = w^{1-\mu} - b/D - \hat{x}t \ , \tag{11}$$

i.e. the real wage, net of resources used per unit housing and commuting from the city edge, equals outside utility. At density D the total population of the city is

$$L = \hat{x}D . \tag{12}$$

Using (12) in (11) gives the labour supply schedule,

$$w^S = \left(u_0 + (b + tL)/D\right)^{1/(1-\mu)}. \tag{13}$$

Comparing this with (6) indicates that the cost schedule has two parts. One is a fixed cost of urban living, which is the extra cost associated with building an urban house, b/D. The other is commuting costs, proportional to transport costs per unit distance and urban population, and inversely proportional to population density D. Equilibrium is, as in figure 2, at the intersection of this labour supply schedule with labour demand.

This structure also provides the basis for a welfare indicator, which contains three parts. One is the real income of workers. In this open city model this is assumed constant and equal to outside utility, u_0. Workers' real incomes are therefore unchanged, independent of city size and other variables that we look at. The second element is rent, accruing to landlords. We assume that this is spent locally on the composite good,[9] i.e. is deflated by price index w^{μ}. The real value of total rent is therefore $w^{-\mu}\int_0^{\hat{x}} R(x)dx$. Integrating, using (10) and (9) together with the condition that $R(\hat{x}) = 0$, the real value of total rent is simply $Dt\hat{x}^2/2 = tL^2/2D$ (the last expression also using (12)).

The third element of welfare is the consumer surplus derived from non-tradables by individuals outside the city. The price of non-tradables, w, affect this consumer surplus, and this is captured by indirect utility function $H(w)$, where, by Roy's identity, quantity demanded is $H'(w) \equiv -h(w)$, as in section 2. Combining these elements, the welfare indicator is

$$U = tL^2/2D + H(w). \tag{14}$$

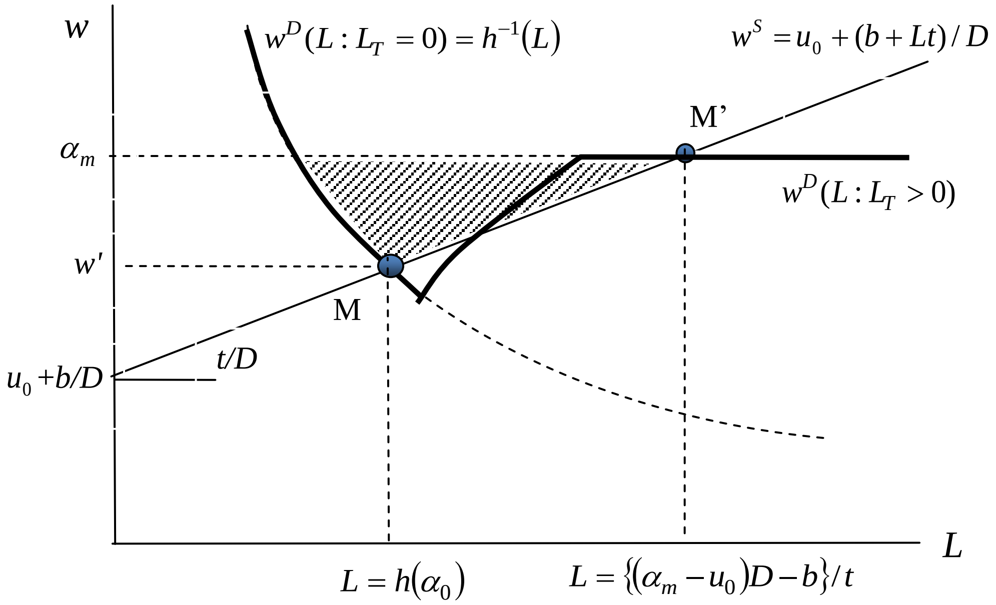
This is illustrated on figure 3 drawn, for simplicity, with $\mu = 0$. Land rent, $tL^2/2D$, is the triangle above the labour supply curve and up to the equilibrium wage, as usual in the urban model. Utility of outsiders is the area under the demand curve for labour in the non-traded sector, down to the wage rate, an area that is smaller the higher is w.[10] The shaded area is therefore the welfare difference between equilibrium M and M'. Higher productivity raises city size and land rent although it also, by raising prices, reduces utility of consumers of non-tradables. The net effect is unambiguously positive.

---

[9] This assumption is necessary for consistency with equation (3), in which fraction μ of income generated in the city is spent on non-tradables.

[10] If $\mu > 0$ so some of the non-traded good is consumed by urban households, then the welfare of outsiders is a fraction of this area.

Figure 3: Welfare gains (with μ = 0).



5. Sunk construction costs and expectations

Costs incurred in constructing buildings are largely sunk, so decisions involve forward looking expectations. In this section and the next we explore the dependence of urban form on these expectations and show how expectations about the future development of the city can be self-fulfilling, this again giving rise to multiple equilibria. Low expectations can be self-fulfilling as the city is constructed at a scale and density that precludes attracting tradable sector production. High expectations can be self-fulfilling as they support construction at a greater scale and density, an urban form that lowers costs, attracts tradables, and generates an outcome with higher welfare.

To capture this we work with two time periods, denoted by subscripts 1, 2, with construction taking place at the beginning of each period. We look at equilibrium in each period, and then tie the two together.

First period equilibrium: The key ingredient is that construction involves sunk costs so first period construction decisions depend on expected prices in period 2, which we denote, for houses and labour, $p_2^E(x), w_2^E$. House prices are expected to satisfy indifference condition (9) in both periods, so

$$p_1(x) = w_1 - (u_0 + xt)w_1^{\mu}, \qquad p_2^E(x) = w_2^E - (u_0 + xt)(w_2^E)^{\mu}. \tag{9'}$$

The expected present value of construction to develop land at x in period 1 is therefore

$$R_1^E(x) = \{\delta p_1(x) + (1-\delta) p_2^E(x)\}D - bw_1^{\mu} \ . \tag{10'}$$

This is analogous to (10), with the weights δ and (1 - δ) capturing the discount rate and length of periods. Land is developed up to the city edge $\hat{x}_1$ (the point at which expected rent is zero) and this gives the total urban population, i.e.

$$R_1^E(\hat{x}_1) = 0, \qquad L_1 = \hat{x}_1 D. \tag{15}$$

Combining (9'), (10'), (15) implicitly defines the first period labour supply schedule,

$$0 = \delta\{w_1 - (u_0 + tL_1/D)w_1^{\mu}\} + (1-\delta)\{w_2^E - (u_0 + tL_1/D)(w_2^E)^{\mu}\} - bw_1^{\mu}/D. \tag{16}$$

Turning to labour demand, we assume that tradable production is not possible in period 1, so the period 1 wage is, from labour demand curve (4a),[11]

$$w_1^D = h^{-1}([1-\mu]L_1). \tag{17}$$

Equations (16) and (17) are the labour supply and demand curves that give the first period equilibrium. Crucially, the first period wage depends on the expected value of the second period wage and, eliminating L_1 from (16) and (17), this relationship will be expressed as $w_1 = \mathrm{W}_1(w_2^E)$. This is monotonically decreasing ; higher expected period 2 wages raise period 2 house prices (9'), and hence period 1 expected present value rents (10'), increasing the built area and city population (15). This increases the supply of non-tradables and reduces their price and the period 1 wage (17). In summary:

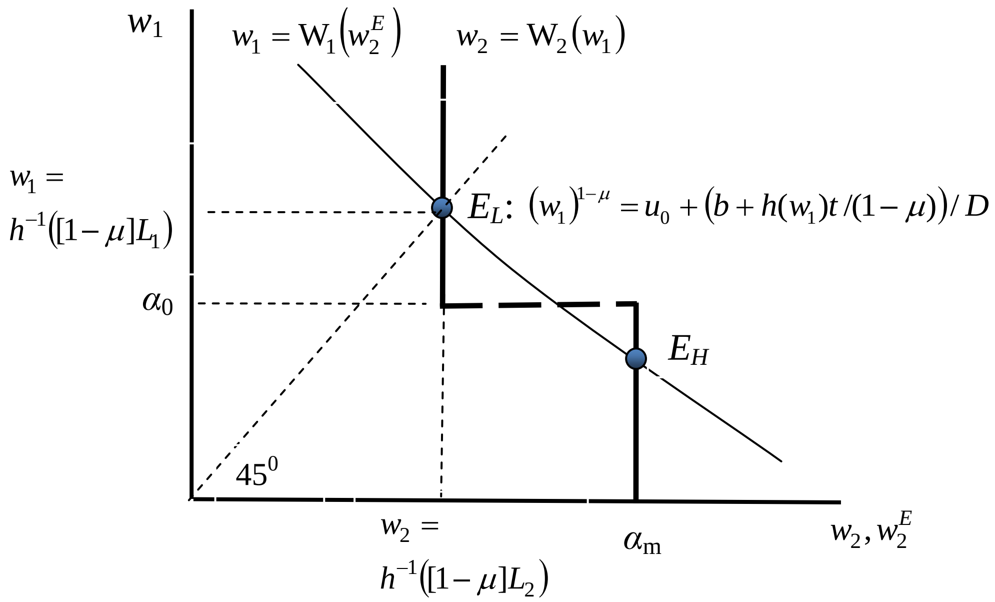
First period equilibrium: $w_1 = \mathrm{W}_1(w_2^E)$, (18)

defined by (16) and (17) and with $dw_1/dw_2^E = \mathrm{W}_1' < 0$.

This relationship is given by the downward sloping line on Figure 4, which has first period wage w_1 on the vertical axis and the second period wage, actual and expected, w_2, w_2^E, on the horizontal. Notice also that if expectations are stationary, so $w_2^E = w_1$, then (16) and (17) give the first period wage defined by $(w_1)^{1-\mu} = u_0 + (b + h(w_1)t/(1-\mu))/D$, as labelled at the intersection of $w_1 = \mathrm{W}_1(w_2^E)$ with the 45° line.

[11] This is a simplifying assumption that avoids the need to look at a range of possible cases.

Figure 4: Multiple equilibria with sunk costs



Second period equilibrium: In the second period house prices, land rent, city edge, and population are given by period 2 variants of (9), (10), (11) and (12), so

$$p_2(x) = w_2 - (u_0 + xt)(w_2)^{\mu}, \qquad R_2(x) = p_2(x)D - bw_2^{\ \mu},$$

$$R_2(\hat{x}_2) = 0 \qquad L_2 = \hat{x}_2 D, \text{ (if } \hat{x}_2 \geq \hat{x}_1\text{)},$$

and hence labour supply curve analogous to (13),

$$w_2^S = (u_0 + (b + tL_2)/D)^{1/(1-\mu)}. \tag{18}$$

Labour demand depends on whether or not tradable production takes place in the city. We suppose that this depends on whether or not the period 1 wage is low enough to trigger entry of tradables, i.e. is less than or equal to α_0. If so, then $w_2 = \alpha_m$. If not, $w_2 = h^{-1}([1-\mu]L_2)$, (the second period version of the labour demand curve absent tradable production, equation 4a). Second period labour demand and wage therefore depends on the first period wage, a relationship that we write as $w_2 = \mathrm{W}_2(w_1)$:

Second period equilibrium: $w_2 = \mathrm{W}_2(w_1)$, taking the form: (20)

If $w_1 \leq \alpha_0$ then $w_2 = \alpha_m$.

If $w_1 > \alpha_0$ then $w_2 = h^{-1}([1-\mu]L_2)$ and, using this in (18) to eliminate L_2,

$$(w_2)^{1-\mu} = \{u_0 + (b + h(w_2)t/(1-\mu))/D\}.$$

This relationship is the step function on figure 4. The two vertical segments are as indicated in equations (20) and the switch point between vertical segments is at $w_1 = \alpha_0$. As illustrated, there are two equilibria. Point E_L has low and stationary expectations, with the city only producing non-tradables and the wage in both periods satisfying $(w)^{1-\mu} = \{u_0 + (b + h(w)t/(1-\mu))/D\}$. Given stationary parameters, this is a solution of both (18) and (20) with $w_1 = w_2^E = w_2 > \alpha_0$. Equilibrium E_H has high expectations of tradable production and high wages in the second period, i.e. $w_2 = \alpha_m$, $w_1 = \mathrm{W}_1(\alpha_m) < \alpha_0$. The high second period wage is perfectly foreseen and leads to more construction and larger first period city size; this in turn reduces urban costs and wages, triggering tradable production.

E_L exists if the wage at this point is greater than α_0. E_H exists if $\mathrm{W}_1(\alpha_m)$ is less than α_0, so that expectation of high second period wages, α_m, lowers nominal first period wages sufficiently to trigger non-tradable production. This is the configuration illustrated in the figure.

6. Building height and urban density

In order to draw out implications more fully we now present results from numerical simulation of the preceding model with the added feature that urban density – the height and form of buildings that are constructed – is endogenous. The previous section assumed that developed areas are built to the same density (i.e. number of houses per unit area) but this underestimates the role of expectations. An important legacy of poor expectations – or other obstacles to urban construction – is that cities are built with low height and density, this amplifying the potential for lock-in to a low-level equilibrium.

To capture endogenous choice of density we suppose that the cost of building density D is a choice variable. The cost of construction is $bw^{\mu}D^{\beta}$, where β is the elasticity of costs with respect to density, $\beta > 1$. Thus, the construction costs of each house built are greater, the greater the density (e.g. height) of development. At each location and each date density is chosen to maximise land rent, and we assume that redevelopment (i.e. demolition of first period construction and rebuilding in period 2) does not occur. In the first period density is chosen to maximise $R_1^E(x) = \{\delta p_1(x) + (1-\delta)p_2^E(x)\}D - bw_1^{\mu}D^{\beta}$ (equation 10') and in the second to maximise $R_2(x) = p_2(x)D - bw_2^{\mu}D^{\beta}$. The first order conditions imply that density is increasing with price, giving a downward sloping density gradient from city centre to edge, $D_1(x),\ D_2(x)$, as expected. The rest of the model is unchanged, except that urban population (12) is now the integrals $L_1 = \int_0^{\hat{x}_1} D_1(x)dx$, $L_2 = L_1 + \int_{\hat{x}_1}^{\hat{x}_2} D(x)dx$, and simulations are undertaken with extra-urban land rent strictly positive, $r_0 > 0$, so the city edge in each period is given by $R_1^E(\hat{x}_1) = r_0$, $R_2(\hat{x}_2) = r_0$. In the simulations reported outside utility is set at unity, $u_0 = 1$, and the two time-periods have equal weight, δ = ½. Values of other parameters are given in the appendix.

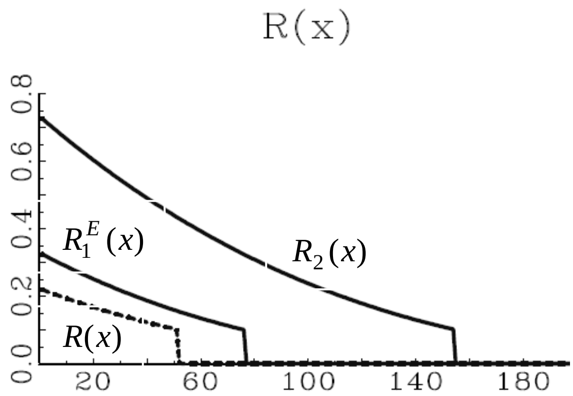
Outcomes are illustrated in figure 5. Equilibrium *L* is that in which no tradable production takes place in either period and stationary expectations are self-fulfilling. The wage in each period is $w = 1.65$, with the 65% premium over outside utility going to meet the urban cost of living, including commuting and (incremental) housing costs. This wage is consistent with no tradable production occurring if $1.65 > \alpha_0$. Intra-city profiles of rent, house prices, and density are given in the three plots, which have distance from the CBD on the horizontal axis. The dashed lines (without subscripts) refer to equilibrium *L*. The city edge is the kink in the rent function (up to which $R(\hat{x}) > r_0$, and beyond which land is undeveloped). House prices fall linearly, since commuting costs are assumed linear in distance. Density and rents decline with distance, as expected.

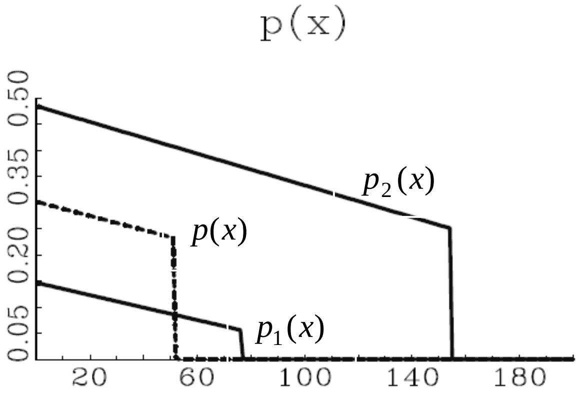
The solid lines, subscripted by period, are equilibrium *H* in which there is tradable production in the second period, but not the first. By assumption, $\alpha_m = 2$, and this is the second period wage. This generates high second period house prices and land rents, and also therefore relatively high expected returns to period 1 building, $R_1^E(x)$. The city is therefore built both larger and denser in the first period than is the case in equilibrium *L*. Period 2 land rents are higher again, so that there is further construction and this is at high density, curve $D_2(x)$; (the upwards step in the density function is an artefact of a two period model, see Henderson, Regan and Venables 2016 for a model in continuous time). The large amount of building in period 1 means low period 1 house prices (recall that construction takes place at the beginning of each period). As a consequence period 1 urban population is larger and the wage lower than in equilibrium *L*. If the wage is low enough ($1.34 < \alpha_0$ in this example) tradable production is triggered and the configuration illustrated is an equilibrium.

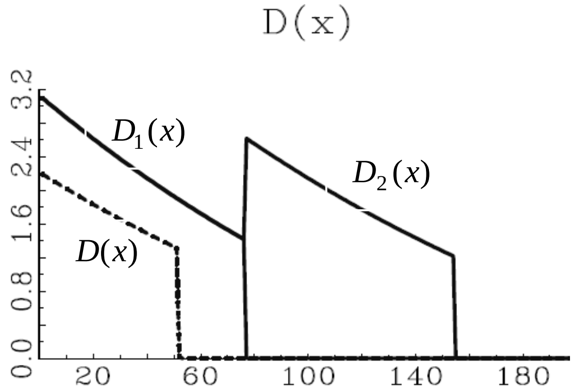
The table in figure 5 additionally gives population and real incomes. Equilibrium *H* supports nearly twice the first period population of equilibrium *L*, and more than three times the second period population. Real income gains accrue in the form of land rents, and the present value of these are nearly four times larger in equilibrium *H* than *L* (and are also larger relative to the wage bill). The growth of the city and expansion of tradable production is associated with varying supply of non-tradables to the hinterland, this increasing hinterland consumer surplus in period 1 but reducing it in period 2. The overall welfare differential between equilibria has present value equal to 20% of the present value of equilibrium *L* city earnings.[12]

Figure 5: Multiple equilibria, expectations and urban density

Equilibrium *L*	Equilibrium *H*	
period 1, 2	period 1	period 2
w = 1.65	w_1 = 1.34	w_2 = 2.0
L = 88.4	L_1 = 167	L_2 = 313
R = 7.9	R_1 = 15.1	R_2 = 54
H = 29	H_1 = 44.8	H_2 = 20
	ΔU_1=23	ΔU_2=37.1







The effects of varying parameters of the model are as would be expected. The first period nominal wage is higher in both the equilibria (i.e. the wage is more likely to exceed α_0, so the economy is less likely to be competitive) the higher are construction costs, *b* and *β,* the higher are commuting costs, *t*, and the greater are *μ* and *h*(*w*), giving city and hinterland demand for non-tradables. The first three of these work by reducing supply of non-tradables, and the last two by raising demand. In all cases the effect is to raise wages and make the city less competitive. The importance of future expectations can be seen by changing *δ*, the weight on the first period relative to the second. A higher *δ* has no effect on equilibrium *L*, and moves the first period equilibrium *H* towards equilibrium *L*, again raising w_1 and making initiation of tradable production less likely.

[12] Equilibrium 1 city earnings have present value wL = 146, and the present value of the welfare change is $(\Delta U_1 + \Delta U_2)/2$ = 30, (with δ = 1/2).

7. Concluding comments

Many cities in developing economies have multiple imperfections. This model has abstracted from these, looking just at the consequences of the interplay between urban costs, tradable and non-tradable production, and expectations. How might other commonly observed imperfections interact with the analysis of this paper?

One set of issues arises in the land market and consequent incentives to build. Often land tenure is insecure and building regulations inappropriately tight (see Collier and Venables 2015). These create powerful disincentives to invest in formal structures and the response is often informal settlement. This delivers low quality housing for which a compensating differential (i.e. a higher nominal wage) is required. Informal settlement also lacks height, a necessary condition for delivering decent accommodation at high density. Matters are further exacerbated by failures in associated markets, in particular the mortgage market may be poorly developed, raising the cost of capital. Each of these factors is a source of inefficiency which necessarily raises urban costs. They could be represented in the model as an increase in building costs, increasing the likelihood that cities will be in the low-level equilibrium.

Weakness of infrastructure provision is widespread (Foster and Briceno-Garmendia 2010). This shows up in congestion and high transport costs, but also affects provision of utilities and public services. These failures affect firms directly as well as raising urban costs for households. They contribute further to making cities less attractive locations for footloose tradable production. Infrastructure investment also plays a role in shaping and coordinating expectations. Commitment to the development of a particular neighbourhood can create positive expectations and a building boom. Without it, uncertainty about where – if anywhere – growth in the city is likely to occur will retard investment. In terms of the model, these inefficiencies raise transport costs, t, and reduce the incentives to build.

Urban public finance also needs to be added to the picture. The high equilibrium creates a tax base (particularly in the form of land value appreciation, see figure 5) which can be used to fund infrastructure, both from current revenues and from borrowing against expected future revenues. In the low equilibrium this ability to fund infrastructure is much reduced. This is a positive feedback mechanism which further reinforces the possibility of multiple equilibria of the type analysed in this paper, and which increases the likelihood of being trapped in the low level equilibrium.

Finally, we have assumed that worker utility is equalised between the city and the hinterland from which it draws population, so there is no direct welfare gain from the movement of workers. Clearly, a gap in favour of the city amplifies the welfare effects that we have described. The presence of such a gap is widely reported, although the full welfare effects have to be placed in the context of the extensive literature following from the Harris-Todaro model (e.g. Williamson 1988).

Capturing all these effects in one model would have led to loss of clarity. However, it seems likely that they all reinforce the risks that cities face, and the possibility that they may follow development paths that exclude them from moving into high productivity tradable goods sectors.

Appendix:

Section 3 and figure 2:

If urban costs are linear, $c(L) = \gamma_0 + \gamma L$, then values of γ at which labour supply passes through the lower and upper kinks on the labour demand schedule are respectively:

$$\gamma_L = (1-\mu)\frac{\alpha_0^{1-\mu} - u_0 - \gamma_0}{h(\alpha_0)}, \quad \gamma_U = (1-\mu)\frac{\alpha_m^{1-\mu} - u_0 - \gamma_0}{h(\alpha_m) + (\alpha_m - \alpha_0)/\alpha}.$$

A region of multiple equilibria exist if $\gamma_U > \gamma_L$, and this inequality is more likely to be satisfied the larger is α; as $\alpha \rightarrow \infty$ a sufficient condition is that $\alpha_m > \alpha_0$. The table below summarises the equations giving wages and city size in each of the four cases illustrated in figure 2, together with the necessary and sufficient condition on the parameter γ for each {w, L} pair to be either type S or M.

Equilibrium	Wage, Labour demand, Labour supply	N & S condition
S	$w = h^{-1}([1-\mu]L) = (u_0 + c(L))^{1/(1-\mu)}$	$\gamma > \gamma_U$
M	$w = h^{-1}([1-\mu]L) = (u_0 + c(L))^{1/(1-\mu)}$	$\gamma_U > \gamma > \gamma_L$
M’	$w = \alpha_m = (u_0 + c(L))^{1/(1-\mu)}$	$\gamma_U > \gamma > \gamma_L$
S’	$w = \alpha_m = (u_0 + c(L))^{1/(1-\mu)}$	$\gamma_L > \gamma$

Section 6: Values used in figure 5 (newhigh.prg)

$u_0 = 1$; $\alpha_m = 2$; $\mu = 0.6$; $t = 0.001$; $\delta = 0.5$; $b = 0.1$; $\beta = 1.5$.

$h(w) = Ew^{-\varepsilon}$, $H(w) = Ew^{1-\varepsilon}/(\varepsilon-1)$, $E = 160$, $\varepsilon = 3$.

References:

Bernard, L., O. D'Aoust, and P. Jones (2016). *The Urban Premium in Africa: Is it Real?* Mimeo, University of Oxford.

Collier,P. and A.J. Venables 'Housing and urbanization in Africa: unleashing a formal market process', in E. Glaeser and A. Joshi-Ghani (eds) (2015), *The urban imperative; towards competitive cities*, OUP, Oxford.

Duranton G. and D. Puga (2005) 'From sectoral to functional urban specialisation', *Journal of Urban Economics*, 57: 343–370

Foster, V. and C. Briceno-Garmendia (2010) '*Africa's infrastructure; a time for transformation*', World Bank, Washington DC.

Gollin, D., R. Jedwab and D. Vollrath 2015, 'Urbanization with and without Industrialisation', *Journal of Economic Growth*, forthcoming

Henderson, J.V. and A .J. Venables (2009) 'The dynamics of city formation', *Review of Economic Dynamics*,12, 233-254.

Henderson, J.V., T. Regan and A .J. Venables (2016) 'Building the city; sunk capital and sequencing', *in progress*

Henderson, J.V., A. Storeygard and U. Deichman, (2015) 'Is climate change driving urbanisation in Africa' processed, LSE

Jones, P. (2016) 'Closed for Business; high urban costs', processed, Dept of Economics, Oxford.

Murphy, K., A. Shleifer and R. Vishny (1989) "Industrialization and the Big Push" *Journal of Political Economy*, 1003-1026.

Nakamura, S., Harati, R., Lall, S.V., Dikhanov, Y.M., Hamadeh, N., Oliver, W.V., Rissanen, M.O., & Yamanaka, M. (2016). *Is Living in African Cities Expensive?* Mimeo, World Bank.

Williamson, J.G. (1988) 'Migration and Urbanisation', in *Handbook of Development Economics*, vol 1, pp425-465, eds H. Chenery and T.N. Srinavasan, North Holland, Amsterdam.

World Bank (2009) '*Reshaping Economic Geography; World Development Report 2009*', World Bank, Washington DC