

25/5/05:

Cities in the developing world*

H.G. Overman, *London School of Economics and CEPR*

Anthony J. Venables, *London School of Economics and CEPR*

Abstract

Rapid urbanisation is a major feature of developing countries. Some 2 billion more people are likely to become city residents in the next 30 years, yet urbanisation has received little attention in the modern development economics literature. This paper reviews theoretical and empirical work on the determinants and effects of urbanisation. This suggests that there are substantial productivity benefits from cities, although unregulated outcomes may well lead to excessive primacy as externalities and coordination failures inhibit decentralisation of economic activity. Policy should operate both by identifying and addressing these market failures, and by seeking to remove institutional obstacles to decentralisation.

Keywords: Urbanisation, economic development

JEL classification: O18, R0

* This work was supported by the UK Department for International Development but does not necessarily reflect DFID views. Thanks to Susan Loughhead and Rachel Phillipson for helpful comments.

Authors' addresses:

H.G. Overman
Dept of Geography
LSE
Houghton Street
London WC2A 2AE, UK

h.g.overman@lse.ac.uk
<http://cep.lse.ac.uk/~overman>

A.J. Venables
Dept of Economics
LSE
Houghton Street
London WC2A 2AE, UK

a.j.venables@lse.ac.uk
<http://econ.lse.ac.uk/staff/ajv>

1. Introduction

The urban population of the developing world is projected to increase by some 2 billion in the next 30 years. Urbanisation rates are strongly correlated with per capita income, productivity tends to be high in cities, and urban job creation is an important driver of economic growth. But urbanisation is also one aspect of the widening spatial disparities that often accompany economic development, and many countries have urban structures dominated by their prime city. While cities are highly productive, they create heavy demands for investments in infrastructure and accommodation, in the absence of which slums and informal settlements develop. Urbanisation gives rise to numerous policy challenges both to make cities work better and to ensure that the overall city structure (the number and size distribution of cities) is as efficient as possible. Cities require major public investments. Some of these are in education, health and other public expenditures that are required in all locations, but others are for reasons specific to urban infrastructure. Furthermore, there is no presumption that an unregulated free market pattern of urban development is socially efficient (even conditional upon appropriate levels of public investment). Urban activity gives rise to many externalities, both positive and negative, so economic theory tells us that an unregulated outcome is inefficient. We observe the grim conditions of developing mega-cities, and we know that in some developing countries the primate city takes a far larger share of population than was the case in much of the developed world at similar stages of development, (see Bairoch 1988 for an in-depth comparison of historical trends). The performance of the urban sector also bears on overall economic growth. Much job creation – both in modern sector activities and in the informal sector – takes place in cities. What determines the attractiveness of a location as a host for investment, and how can city environments be developed to maximise job creation? Do ‘bad’ city structures impede overall growth?

These concerns point to the need to have a robust understanding of the economics of cities, both theoretically and empirically. Unfortunately, such an understanding does not exist. Work in development economics over the last several decades has been almost silent on the issue. Some aspects of urbanisation have been approached in the contexts of migration and industrial development, but rarely focussing on the particularly urban issues that arise. Urban

economics is, however, having a renaissance in the academic literature.¹ There has been improvement in theoretical methods, as economic analysis steadily comes to do a better job at analysing economies of scale, and there have been empirical advances as larger data sets and better econometric tools are applied. Some of this work has been applied to developing countries (most notably by Vernon Henderson from Brown University), although the developing country literature remains thin.

The objective of this paper is to draw out the messages from this literature (analytical, and empirical, developed and developing country) that can inform thinking about cities in developing countries. We seek to understand several broad issues. The first is the benefits and the costs of cities. Per capita income and productivity is generally higher in cities than elsewhere, partly because of productivity benefits of cities, and partly because of political access and rent seeking. At the same time urban living is constrained by land availability and brings congestion and pollution costs. How do these costs and benefits depend on urban scale, and how are they shaped by urban governance? The second broad issue is the extent to which the private costs and benefits of cities diverge from social ones. We will argue that cities are riddled with market failures. These cause inefficiencies within cities, distort the city structure (the size and number of cities) and possibly also reduce the returns to job creation, impeding overall growth. Third, we offer some remarks about policy. Researchers' identification and quantification of particular sources of market failure and externality is generally not good enough to apply detailed policy targeting. However, we argue that there is a tendency for cities to be too large (and for there to be too few of them), and that policy should seek to remove obstacles to decentralisation.

We organise the material by first looking at the analytical work on costs and benefits of cities and associated market failures, and then turning to empirical studies. Section 2 presents an overview of analytical approaches to urban development, and we will see that alternative theories give different predictions as to the real income benefits or costs of cities. In section 3 we identify the main market failures associated with cities and the extent to which they lead cities to be larger or smaller than is efficient. We then turn to the evidence, looking at both outcomes and the underlying mechanisms that support these outcomes. In section 4 we look at evidence on

¹ See for example the new Handbook of Regional and Urban Economics (Henderson and Thisse 2004). Not one of the twenty chapters in this volume deals with developing country issues.

mechanisms. What evidence is there on the relationship between productivity and city size? Are there increasing returns to infrastructure investments, making it efficient to concentrate such investments in cities? Alternatively, are cities supported by rent seeking forces and access to the political process? Section 5 turns from mechanisms to outcomes. What do we know about determinants of the urban size structure, and how does it change over time? How does urban structure itself bear on the overall performance of an economy? Finally, in section 6, we conclude with a discussion of policy issues.

Before embarking on the literature, it is helpful to have a brief review of facts. The basic numbers on urban populations and their possible development over the next decade are given in table 1. The total population in cities with population of greater than 1 million is predicted to increase, between 2000 and 2015, by 5% in Europe and 19% in North America. In contrast, in LAC the predicted increase is 46%, in Asia it is 56%, and in Africa 96%, this following a near five-fold increase between 1975 and 2000. During the development process, population becomes highly concentrated in ‘primate’ cities. However, there is considerable evidence that the degree of primacy follows an inverse U-shape, as urban population – and modern sector activity more generally – disperses in more mature economies (see section 5.1 below).

Cities offer higher income levels than average in the country. The relationship between cities and productivity is well researched for developed countries (see section 4.1), and some numbers for developing countries are given in table 2. This is based on sample data compiled by the UN-Habitat programme. We see that per capita income in cities in Africa is 65% higher than the national average. It is also the case that urban poverty rates are generally lower than rural rates (Kessides, 2004). Yet at the same time cities support extremely high levels of both unemployment and informal sector activity.

Finally, the creation of new jobs is overwhelmingly urban. While data on job creation by location is hard to obtain, GDP growth by sector indicates that, for some regions, virtually all job creation taking place in activities that are generally urban based.

TABLE 1. DISTRIBUTION OF THE TOTAL POPULATION OF MAJOR AREAS BY TYPE OF SETTLEMENT AND SIZE OF URBAN SETTLEMENT, 1975, 2000 AND 2015

Major area	Type of settlement and number of inhabitants of urban settlement	Total population (millions)			Percentage distribution			Growth rate (percentage)	
		1975	2000	2015	1975	2000	2015	1975-2000	2000-15
Africa	Total population	406	794	1 110	100.0	100.0	100.0	2.68	2.24
	Urban Population	102	295	503	25.2	37.2	45.3	4.23	3.56
	10 million or more	0	0	27	0.0	0.0	2.5
	5 million to 10 million	6	23	21	1.5	2.9	1.9	5.35	-0.63
	1 million to 5 million	12	64	123	3.0	8.1	11.1	6.67	4.35
	500,000 to 1 million	14	26	43	3.5	3.3	3.9	2.50	3.31
	Fewer than 500,000	70	181	288	17.3	22.9	25.9	3.80	3.08
	Rural areas	304	498	607	74.8	62.8	54.7	1.98	1.31
Asia.....	Total population	2,397	3,672	4,371	100.0	100.0	100.0	1.71	1.16
	Urban Population	592	1 376	2 005	24.7	37.5	45.9	3.37	2.51
	10 million or more	31	136	214	1.3	3.7	4.9	5.90	3.01
	5 million to 10 million	47	87	162	1.9	2.4	3.7	2.50	4.12
	1 million to 5 million	147	318	466	6.2	8.7	10.7	3.08	2.55
	500,000 to 1 million	78	150	193	3.3	4.1	4.4	2.61	1.72
	Fewer than 500,000	289	684	969	12.0	18.6	22.2	3.45	2.32
	Rural areas	1,805	2,297	2,366	75.3	62.5	54.1	0.96	0.20
Europe	Total population	676	727	705	100.0	100.0	100.0	0.29	-0.21
	Urban Population	455	534	538	67.3	73.4	76.3	0.64	0.04
	10 million or more	0	0	0	0.0	0.0	0.0
	5 million to 10 million	37	32	32	5.4	4.4	4.6	-0.52	0.01
	1 million to 5 million	82	110	117	12.2	15.1	16.5	1.15	0.41
	500,000 to 1 million	46	49	48	6.8	6.7	6.8	0.22	-0.14
	Fewer than 500,000	290	344	341	42.9	47.2	48.4	0.67	-0.05
	Rural areas	221	193	167	32.7	26.6	23.7	-0.54	-0.98
Latin America and the Caribbean	Total population	322	519	630	100.0	100.0	100.0	1.91	1.30
	Urban Population	198	391	507	61.4	75.4	80.5	2.73	1.73
	10 million or more	21	59	66	6.5	11.3	10.5	4.11	0.82
	5 million to 10 million	17	20	36	5.3	3.8	5.6	0.56	3.94
	1 million to 5 million	32	86	139	10.1	16.5	22.1	3.90	3.23
	500,000 to 1 million	18	39	45	5.5	7.5	7.2	3.14	1.00
	Fewer than 500,000	109	188	221	34.0	36.3	35.0	2.17	1.07
	Rural areas	124	127	123	38.6	24.6	19.5	0.10	-0.25
Northern America.....	Total population	243	314	356	100.0	100.0	100.0	1.02	0.83
	Urban Population	180	243	288	73.8	77.4	81.1	1.21	1.14
	10 million or more	16	30	32	6.5	9.5	9.1	2.54	0.53
	5 million to 10 million	16	7	13	6.4	2.2	3.7	-3.23	4.28
	1 million to 5 million	52	85	100	21.3	27.0	28.2	1.96	1.14
	500,000 to 1 million	17	26	23	7.0	8.4	6.5	1.73	-0.81
	Fewer than 500,000	79	95	119	32.6	30.3	33.4	0.73	1.50
	Rural areas	64	71	67	26.2	22.6	18.9	0.44	-0.37

Table 2: Income and employment (1996 sample data)

Region	GDP per capita	City product	Informal employment	Unemployment rate
Africa	\$441	\$729	54%	23%
Arab States	\$2,752	\$3,170	65%	11%
Asia Pacific	\$4,742	\$6,182	33%	8%
LAC	\$3,350	\$3,226	39%	13%
Transitional	\$2,541	\$2,905	21%	9%
All developing	\$2,670	\$2,988	37%	12%
HIC	\$22,501	\$22,103	3%	8%

http://www.unhabitat.org/programmes/guo/guo_analysis.asp

2. Urban economics

Two broad theses are offered to explain urbanisation in developing countries, and in particular the role of the primate city. The first, which we will refer to as ‘productivity’, is that there are various efficiency gains associated with clustering activity; firms and workers are attracted by these benefits. The second, which we will refer to as ‘rent seeking’, is that city dwellers have preferential access to the political system, this enabling them to extract rents (relative to non-urban dwellers). Of course, these hypotheses are not mutually exclusive; both operate to varying degrees in different countries and cities. And pulling in the other direction there are costs associated with urban centres, these including transport and infrastructure costs, and externalities associated with congestion and pollution.

2.1 Productivity:

To set out a way of thinking about these issues, note that there are some functions that are inherently urban, or at least appear so in virtually all historical contexts. These include government and central administration. Other functions can potentially operate either in a city or

in smaller towns or villages. These include distribution (markets, exchange, wholesale and retail activities) other services and manufacturing, although the size of a manufacturing plant places a lower limit on the size of a town or village that can accommodate it. Our main question is; what determines the benefits of grouping such activities in an urban centre?²

The first argument derives from ‘thick’ goods markets. A system of exchange works better if it operates at a reasonable scale. The variety of goods on offer is greater, search and travel costs are reduced, and competition is more intense. Unsurprisingly then, ‘market towns’ develop in order to provide locations for trade. Often these centres are home to local agricultural markets and so have important linkages with the surrounding rural economy.³

The second urban advantage derives from ‘market access’. Suppose that a manufacturing (or service) activity faces transport costs on the goods that it sells. Other things being equal, the most profitable location for such a firm is close to a large mass of consumers – i.e. in an urban centre. Models predict an amplification effect, so that manufacturing is drawn more than proportionately towards locations with good market access. Physical geography and transport systems also come into play here; a port or river crossing will have better market access than a mountain-top. These arguments extend to forward and backward linkages (perhaps better labelled as cost and demand linkages)⁴. Many firms are engaged primarily in supplying other firms (rather than final consumers), and for such firms good market access means proximity to customer firms. This is the backward or demand linkage – firms want to locate close to the sources of demand for their output. But the converse of any backward linkage is a forward or cost linkage. Firms that purchase the output of other firms will want to locate close to their supplier firms. The combined effect of these demand and cost linkages can create a powerful force for agglomeration of activity. We see it in dense networks of firms in related industries – engineering, electronics and even financial services, where firms that supply specialist financial skills locate near the big financial institutions, and these institutions benefit from access to the skills of the specialists. A developing country example is the surgical instruments cluster in Sialkot, Pakistan, where Nadvi (1999) identifies “over 20 stages in production, each requiring

² Many of the arguments presented here are developed in more detail in Duranton and Puga (2004).

³ The particular policy issues relating to small and intermediate urban centres and their potential role in regional and urban development are beyond the scope of this paper. They are discussed in depth by Tacoli (2004).

⁴ For a formal analysis see Fujita et al (1999). An older tradition of development economics considered these linkage, although generally without recognising the importance to the argument of increasing returns to scale and associated market failures.

distinct skills and tools. Surgical instrument making thus lends itself to an extensive division of labour and the process of large numbers of subcontractors in most process activities” (Nadvi, 1999, p 87).

Several other arguments also point to productivity benefits of cities. Thick market effects arise in the context of labour markets. Large pools of specialist workers and of the firms that use these skills benefit from better matching of skills with requirements, and also from risk sharing if there are firm or worker specific fluctuations in demand or supply. Incentives to acquire skills are greater if the skills are sought by several firms, so the worker is less likely to be subject to the monopsony power of a single employer. Labour turnover is one – but not the only mechanism – through which firms in a dense cluster of activity can benefit from the skills and knowledge of other firms. There is considerable evidence of productivity spillovers between firms, as they are able to learn about and imitate the practises of other firms in the industry. Silicon valley provides an example where knowledge exchange – formal and informal – is quite widespread. The knowledge may be about production methods, marketing skills, or simply knowledge about the location itself. Thus, multinational firms tend to cluster in particular locations, partly because one firm, observing the success (or failure) of another, learns about the quality of the business environment in the location. Hausman and Rodrik (2002) argue that very narrow patterns of specialisation in developing countries (for example, specialization in soccer ball production) arise as producers learn about the efficiency of a particular location for producing a particular good, this then becoming public knowledge.

The final argument is to do with the provision of public goods. It may be that there are increasing returns in provision of these goods. For example, if a country is to build 100 miles of paved road, it may be efficient to build most of it joined together rather than scattered around. In this case an ‘urban bias’ in public expenditure and provision may be an efficient allocation of resources (Arnott and Gersovitz 1986).

These are distinct arguments, but all share several characteristics in common. First, they are sources of spatially concentrated increasing returns to scale; there are efficiency gains from having things locate in the same place. Second, all these arguments are potential drivers of cumulative causation. People choose to set up activities in a location not because of the intrinsic merit of the location, but because other people have already done so, or are expected to. Put differently, there are positive reciprocal externalities; my presence makes the city more attractive

to you, your presence makes it more attractive to me, and so on. Furthermore, these mechanisms create ‘lock-in’ or ‘path-dependent’ development. A city may be in the wrong place or an industry in the wrong city – but once there, it will not be profitable for any single producer to move away from the cluster.

Before leaving the issue of the productivity benefits of cities, we should note that cities may have dynamic effects over and above the static effects that we have focused on here. This dynamic effect will depend on the role that urban environments play in developing new products and processes. Duranton and Puga (2001) argue that large diverse metropolitan areas play a role as a “nursery”. These information rich environments allow firms to develop new products and processes. However, once production becomes standardised, firms move out to cheaper specialised locations where they can benefit from localisation economies without the high diseconomies of the large urban city. Duranton and Puga (2001) provide evidence on firm births and relocations across French metropolitan areas that appear to be consistent with their theory.

Could large diverse metropolitan areas be playing a similar role in developing countries? Clearly, the kind of R&D and innovation undertaken by developing country entrepreneurs differs from the way these terms are commonly used in the developed country context. Nevertheless, entrepreneurs in low-income countries must also engage in a process of innovation and learning. Their focus is on what Rodrik (2004, p. 9) calls cost discovery: “What is involved is not coming up with new products or processes, but discovering that a certain good, already well established in world markets, can be produced at home at low cost”. Rodrik suggest some developing country examples: cut flowers in Colombia, T-shirts in Bangladesh, soccer balls in Pakistan and software in India. The urban nature of these cost discovery processes remains largely unexplored. However, Hausmann and Rodrik’s (2002) emphasis on the importance of tacit knowledge (the kind that cannot be easily codified in to blueprints) in the self discovery process strikes a chord with urban economists who have long seen such knowledge as playing a key role in the information spillovers that occur within cities. This suggests that, just as for their developed country counterparts, this process of cost discovery is likely to be significantly easier in the information rich environment of large diverse urban areas. Unfortunately, theoretical and empirical work on these issues in a developing country context is not well developed and this is an important area for future work.

2.2 *Rent seeking:*

The arguments above turned on real efficiency gains from the scale effects of urban centres. Other arguments are based on the idea that urban-rural differentials are due to transfer payments – urban dwellers benefit not by creating resources, but by extracting them from the rest of society. These arguments have been developed by a number of researchers, including Lipton's view of 'urban bias' (Lipton 1976, 1993) and Hoselitz's 'parasitic city' (Nash 1977).

The main mechanism here is political access. For example, in many developing countries, starting a business, hiring and firing workers, registering property, enforcing contracts, getting credit, protecting investors and closing a business are subject to extensive regulation (World Bank 2005). If business activity is extensively regulated by permits and licenses, then the probability of getting them may be enhanced by proximity to the administrative centre. Similar arguments may apply not just to firms, but also to households. The political power of an urban proletariat may mean that the government acts to raise their real incomes. This will attract workers to the town (see next sub-section), although its effect on employment is ambiguous; food subsidies may reduce the wages that firms need to pay, creating jobs, while minimum wage legislation will have the opposite effect.

Transfers to urban dwellers can also occur through 'biased' public expenditure programmes, ranging from health through to transport. Such programmes will tend to expand city population although, as we have noted, there may be efficiency gains from concentration of public expenditures in urban areas. The extent of these gains is an empirical matter, and the trade-off between such gains and spatial inequality is a normative concern.

The urban bias argument is also made in terms of overall patterns of import protection and relative prices. Import substituting governments have typically raised the prices of manufactures relative to agricultural goods, and this is sometimes argued to be a source of urban bias. For present purposes however, it is important to keep separate the question of what is produced and where it is produced. Supporting manufactures is supporting cities only if – for some other reason – manufactures are produced in cities.

2.3 *Urban costs:*

The forces outlined above were to do with cities creating income (productivity effects) or transferring it (rent seeking). However, cities also destroy income – they create costs for urban dwellers.

One source of cost is the travel and commuting required by urban inhabitants. A standard urban economics model assumes that jobs are clustered in one (or several) ‘central business districts’ to which workers have to commute. This in turn generates a land rent gradient; rents are high in the centre where commuting costs are low, and low on the edge where commuting costs are high. Commuting costs are the resource and time costs of running an urban transport system, which may be amplified by congestion costs.

Conceptually, the costs outlined in the preceding paragraph are of three quite distinct types. The first is the direct cost of uncongested commuting; a resource cost, but one that is not necessarily associated with any sort of market failure. The second is the cost of paying urban rents. This is a transfer payment, not a resource cost; it is paid by urban dwellers to urban landowners. The third is the addition to commuting costs created by congestion, a negative reciprocal externality between those travelling within the city. Other negative externalities also create costs for urban dwellers, most notably air and water pollution.

A further source of cost, that has been the subject of much attention in the development literature, is that a city may attract a number of workers who are un- or under-employed. The economics underlying this dual labour market structure is illuminated by the Harris-Todaro model. The model supposes that the urban real wage is above real earnings in agriculture. This may be because of institutional rigidities supporting a high urban wage (minimum wage legislation, union power or price support), or for efficiency wage reasons (wage reductions are unprofitable as they reduce the quality of labour working in the firm through nutritional, effort, or selection effects). The high urban wage attracts an inflow labour to the city, in search of these ‘formal sector’ jobs. Given the number of such jobs, equilibrium migration is attained when the probability of a migrant getting a job is low enough that the expected wage from migration equals earnings in agriculture. Migrants who fail to get a formal sector job are unemployed, or work for a much lower wage in the urban ‘informal sector’.⁵ The strength of the Harris-Todaro model is its simplicity, but this of course comes at the cost of abstracting from many important

⁵ See Becker and Morrison (1999) for extensive discussion of rural urban migration.

aspects of the problem. The dual structure of formal and informal urban labour markets is complex, and rural-urban migration occurs for many reasons. Nevertheless, the model makes the point that a possible cost of urbanisation is the associated development of a mass of low wage and low productivity urban labour.

3. Urban market failures

Most of the arguments above are associated with some sort of market failure. Consequently, there is no presumption that an unregulated outcome (even one in which public investments are being made efficiently according to a cost-benefit rule) will be efficient. But what are the key market failures that should be borne in mind in thinking about city growth? In this section we lay out some of the ingredients needed to think about the applied welfare economics of urban growth. They suggest that an unregulated outcome will lead to the largest cities being oversized, essentially because firms are deterred from setting up in smaller cities both because they discount the future growth prospects of the city, and because they do not internalise the externalities they create in such cities.

3.1 Determinants of city size.

Suppose that at least some of the productivity effects of section 2.1 are present, as are increasing urban costs – specifically the commuting costs and urban rents of section 2.3. The (private) profits earned by creating a job in a city are then illustrated by the line $PR(n)$ on figure 1.⁶ Employment in this particular city is denoted by n and measured on the horizontal axis. The curve $PR(n)$ (profit) is the value of output per worker minus the wage rate. It is initially increasing because of the productivity advantages of increasing the scale of urban activity. The rate of increase diminishes and then (possibly, and as illustrated) eventually turns down, because of increasing urban costs for which workers need to be compensated by higher nominal wages. The curve $ASB(n)$ (average social benefit) gives the total real income created by the city, expressed per worker. This is the value of the extra output created, net of costs incurred, in having workers in the city rather than in alternative employment (for example in agriculture).

$MSB(n)$ is the marginal social benefit, derived from $ASB(n)$ and necessarily intersecting it from above at the maximum of $ASB(n)$, point m . The gap between the social and private returns to creating a job, $MSB(n)$ minus $PR(n)$, is the value of the externalities created by an additional job. To the left of point a this externality is positive; that is, the productivity benefits created by an expanding urban employment exceed any external costs (such as congestion) associated with the expansion. To the right population is so large that this is reversed.⁷

The curves on figure 1 describe per worker returns in a single city. Of course, there may be many such cities, each described by a figure such as this, although possibly with different shape curves reflecting both the natural advantages and the differing functional specialisation of each city. Total urban population can therefore change either by moving along the horizontal axis for existing cities, or by growing new cities.

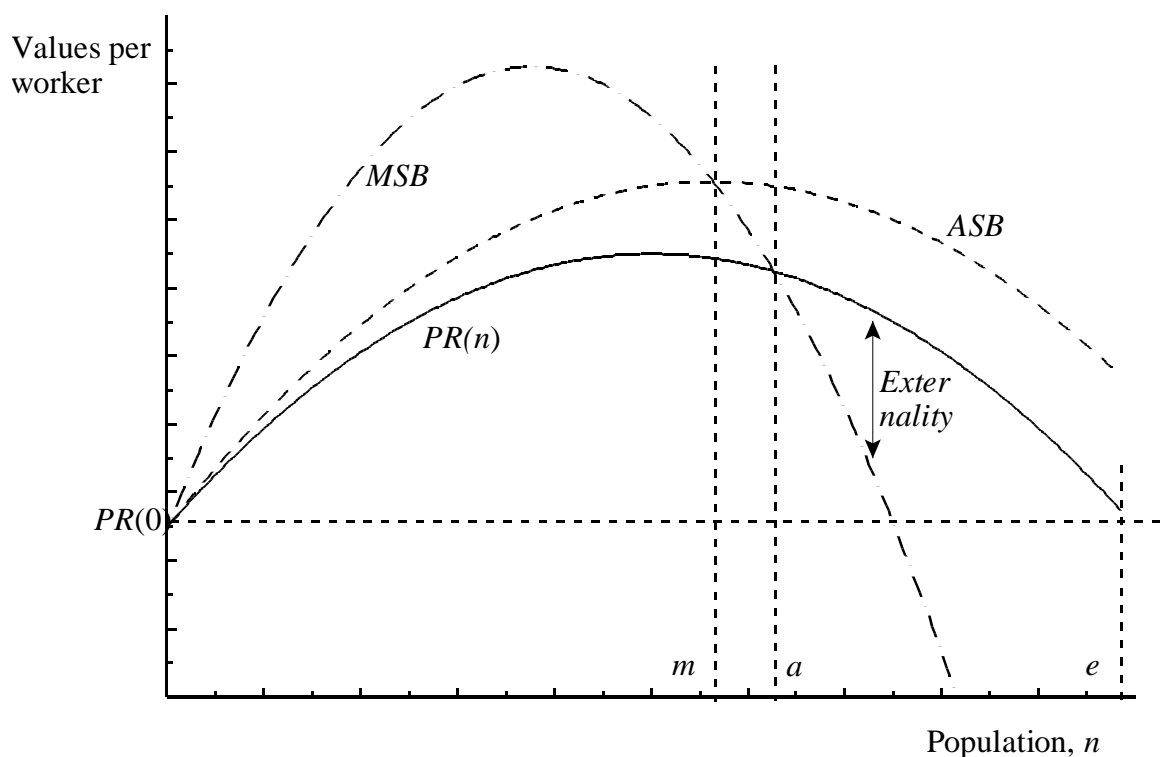


Figure 1: Private and social returns to city size

⁶ For development and extensions of models on which this section is based see Henderson 1974, Helsley and Strange 1991, Black and Henderson 1999, Fujita, Krugman, Venables 1999, Duranton and Puga 2001, Henderson and Venables 2004.

These curves just summarise the basic technology of the way the city operates. What economics do we learn from them? The first point is that as city size changes so do the social returns to creating a job, $MSB(n)$. This means that we can define an optimal city size at which the social benefits are maximised. This is at point m . This point yields highest benefit per worker; it is efficient to grow each city to this point, and then expand overall urban population further by replicating cities rather than expanding this city further.⁸

The second point is that the marginal social return exceeds the private return of job creation, $MSB(n) > PR(n)$, for city sizes up to some point, a , and is then less. As we have seen the difference between these curves is the net externality generated by an additional urban worker. The point is that private incentives to create jobs, as determined by $PR(n)$, are generally different from social incentives.

Given these externalities, what outcomes can occur? One possibility is that a set of institutions develop that internalise the market failures and lead to the efficient outcome. As shown by the work of Henderson (1974), Henderson and Black (1999) this will occur if a single large property developer owns all the land in the city and is able to tax and/or subsidise inhabitants of the city. It also occurs if there is local city government responsive to the incumbent population and with similar tax/ subsidy powers. However, the way these institutions must work, if the outcome is to be efficient, is that competition between cities leads the large developer/ city government to subsidise workers to enter the city, with the cost of the subsidy being recouped through land rent. While this may be a good model of the development of retail malls in the US, these are not institutions and policies that we typically see implemented in urban government, least of all in developing countries.

If left to private market forces, what outcomes do we expect, and how large does each city become? This is quite a difficult question to answer, depending critically on dynamics and the expectations of firms. A naïve first answer is that the city will expand to point e on the figure. The argument is that if a firm sets up in a new location (eg a secondary town) then it will not benefit from the productivity advantages of the cluster, so its profits will be close to $PR(0)$.

⁷ Point m is to the right of the maximum of $PR(n)$, because there are elements of urban costs (transport, rent) that increase with city size, but are not externalities.

⁸ If later cities are somewhat less efficient – eg less natural advantage – and slow to grow, then it will be efficient to expand existing cities somewhat beyond this point.

Point e is constructed such that a firm in a city of size e has the same profits as a firm in a new city, $PR(e) = PR(0)$. Point e is usually, although not necessarily to the right of point m , indicating that the free market outcome produces cities that are larger than is socially efficient.⁹

This answer is somewhat naïve, because it is based on a firm failing to anticipate that the new city to which it moves might grow, moving along and up its $PR(n)$ schedule. At the other extreme we might suppose that the firm (rationally) expects to be followed by other firms, and perfectly anticipates returns moving along the $PR(n)$ schedule.¹⁰ This will foster earlier formation of new cities and move city size towards the socially efficient level. However, it will still not be efficient, because even this calculation is based on the path of private returns, $PR(n)$, rather than the path of social returns, $MSB(n)$. It will tend to be the case that cities will be smaller than socially optimal if positive externalities are increasing over a very wide range of city sizes, and larger than socially optimal if these externalities turn down (and possibly become negative).¹¹

The conclusion so far is that, despite ambiguities turning on both the behaviour of firms and the precise nature of externalities, there is a likelihood that unregulated cities will be larger than socially optimal. Essentially, firms are reluctant to move to a new city (relative to the efficient outcome), both because they might discount the future growth prospects of such a city, and because they do not internalise the positive externalities that they create.

3.2 Other market failures

The discussion above was couched in terms of productivity effects versus the costs of urban scale, and the extent to which these benefits and costs were associated with externalities. How are the arguments modified if rent seeking behaviour or under-employed labour enters the picture?

The effects of rent seeking behaviour turn on who gets the rents. One possibility is that they accrue to the owners of firms in the primate city (via access to permits etc). This provides an additional reason why it is difficult to establish firms in a new location, reinforcing the advantage of the primate city and increasing the likelihood that it is larger than would be efficient. (In terms of figure 1 the $PR(n)$ schedule is shifted up for the primate city, but not for

⁹ It is possible that point e is to left of point m , see Venables (2005).

¹⁰ It may also need to borrow on efficient capital markets to cover short run losses.

other cities. This shifts the point e to the right, since the comparison value, $PR(0)$, is that for other cities).

Alternatively, the rent may accrue to workers in the primate city, in which case firms are induced to leave the primate city in search of lower wage labour elsewhere. However, the rent will attract workers for whom formal sector jobs are not available. This is the core of the Harris-Todaro model, in which such rent induces excess migration. The model contains two market failures. The wage is above the market clearing level, and there is excess migration and consequent urban employment. In the simplest version of this model, it turns out that the relationship between equilibrium and optimum city employment levels is unaffected. The private returns to job creation are reduced by the high wage and the social returns reduced by the additional employment, the two effects exactly cancelling out. (The $PR(n)$, and the $MSB(n)$ schedules shift down by the same amount). However, the city population is increased, because of the presence of the unemployed/ informal sector inhabitants, and we return to discussion of the informal sector in section 4. Further work needs to be done integrating urban economists' models of the city with the migration models of Harris-Todaro et al. For a start on this programme of research see Brueckner (1990) and Brueckner and Zenou (1999).

A number of further market failures are also likely to be important. One that has attracted attention in the literature is the inefficiency of urban land markets. If property rights over urban land are poorly defined then its use will be inefficient, and this may reduce urban densities, in turn increasing urban space requirements and commuting costs. Thus, Maxwell (1995) reports that in 1992 56% of land in Kampala was devoted to agriculture, creating significant real income loss (see also Becker and Morrison 1999). Investment climate surveys find that, in some countries, access to urban land (including costs of and barriers to buying and selling land) is a major obstacle to business activity (Kessides 2004).

3.3 City size, job creation and growth

Our attention so far has been primarily on the distribution of jobs between cities, given that the jobs are being created. But a further implication of the analysis above is that the private returns to job creation are likely to be below the social returns, and certainly below the social returns that would accrue if cities were optimally sized. This means that the presence of urban market

¹¹ See Henderson and Venables (2004).

failures and associated inefficient city structures has the effect of depressing the returns to job creation. If job creation is itself endogenous, this will slow down job creation and overall economic growth (see Henderson 2004).

4. Evidence: benefits and costs of urban size

In figure 1, the costs and benefits of cities changed with city size so that the average social benefit peaked at an “optimal” city size (m). In addition, the existence of externalities means that the marginal social benefit $MSB(n)$ does not coincide with the marginal private benefit $PR(n)$ so that a free market economy will not necessarily reach this optimal city size. In this section we consider two questions. First, what do we know about the existence and magnitude of externalities? Second how do the costs and benefits of cities change with city size? We will start our review of the empirical literature by looking at productivity effects. That is, we will see whether city size and structure impact on firm productivity. The evidence shows that concentrations of economic activity do generally increase firm productivity, although different types of activity may be affected in different ways. These productivity effects allow firms to pay higher wages, but the benefits to workers are offset by higher cost of living. After reviewing the evidence on productivity effects, we will briefly consider what the literature has to say about these cost of living effects. Finally, we will consider the literature that puts the two offsetting forces together to consider how real income changes with city size. It turns out that the amount we know about each of these different effects and their relationship to city size decreases as we work through the list. Thus, there is a large literature that attempts to identify the nature of production externalities – i.e. the impact of urban characteristics on firm productivity. There is surprisingly little evidence on cost of living effects as a function of city size. Finally we know very little about the relationship between pollution and city size and even less on the sources of positive non-production externalities. Let us now work through these questions in turn.

4.1 Productivity

The main areas of enquiry revolve around the scope, sources, and magnitudes of productivity effects (Rosenthal and Strange, 2004). There has been relatively little consideration of the first two of these. For geographical scope, data requirements mean that analysis is usually some

broadly defined metropolitan statistical area. Similarly, there is little evidence on the temporal dimension with most authors focusing only on contemporary effects. With regard to industrial scope, the issue is whether interaction is greater between firms within a particular sector, or across sectors. Empirical work has usually resorted to a dichotomy and distinguished between two types of externalities: localisation and urbanisation. Localisation economies exist if firms benefit from the presence of firms in the same industry. Urbanisation economies exist if the benefit arises from just being in large urban areas. Some authors, but not all, define urbanisation economies as arising from large *diverse* urban environments.

Several stylised facts emerge from the developed country literature. There is consistent evidence that productivity increases with city size. The survey by Rosenthal and Strange (2004, p. 2133) suggests that “[i]n sum, doubling city size seems to increase productivity by an amount that ranges from 3-8%”. When studies attempt to distinguish between urbanisation and localisation economies the strength of these economies can vary substantially across industries. For example, higher tech and service activities appear to benefit from urbanisation economies while more standardised production appears to benefit from localisation economies. Within the standardised production activities, some industries show very strong localisation externalities, while the affects for other industries are much weaker. Generally the most robust findings concern the existence of localisation economies.

Most of the econometric evidence that is available refers to developed countries but, as we stated above, there is nascent literature on productivity effects in developing countries. Table 3 presents an overview of the available papers and highlights their findings. The econometric literature summarised in Table 3 is still in its infancy. Our reading is that, with the exception of two of the studies on India, the findings for developing countries are broadly in line with those from developed countries. Unfortunately, in the particular case of India, the finding of localisation diseconomies in some sectors sits uneasily with that of significant spatial concentrations of particular industries reported in Lall et al (2004). Hopefully, further work on the detailed Indian data used in Lall et al (2003, 2004) will throw further light on this matter. Results for Brazil, China, Indonesia and Korea suggest that there is some evidence of urbanisation economies, but not in all industries. Interestingly, little of the developed or developing country literature considers how these externalities change with city size. In fact, nearly all specifications are log-linear which implies constant elasticity of productivity with

respect to own industry concentration or diversity. That is, the externalities created per worker are basically independent of city size.

Table 3: Developing Country Evidence on Productivity Effects

Country	Paper	Findings
Brazil	Henderson (1988)	Localisation
China	Chen (1996)	Localisation economies for 2 out of 2 industries (machinery and food); does not consider urbanisation economies
India	Shukla (1996)	Localisation and urbanisation economies; urbanisation stronger than localisation in 11 out of 13 industries
	Mitra (2000)	Urbanisation economies in 11 out of 17 industries; does not consider localisation economies.
	Lall et al. (2003)	Urbanisation: Food processing, textiles, leather, paper, chemical, basic metals, mechanical machinery, electrical Localisation diseconomies
	Lall et al. (2004)	No localisation or urbanisation
Indonesia	Henderson and Kuncoro (1996)	Localisation: Apparel (inc. textiles), non-metallic minerals, machinery (inc transport and electrical) Urbanisation: Wood, furniture, publishing
Korea	Lee and Zang (1998)	Localisation not urbanisation economies (19 industries)
	Henderson et al (2001)	Localisation: Traditional, heavy, transport, machinery Urbanisation: High tech

The extent of localisation economies can vary substantially across industries. Figures reported in table 4 for Korea give some idea of this variation. The second column shows the strength of localisation economies in different industries. The coefficients give the elasticity of productivity in an industry with respect to own industry employment in the same metropolitan area. Thus, for heavy industry a 1% increase in own industry employment increases productivity by 0.08 percent. While this might not appear to be a large effect, the implication is that moving from a city with employment of 1,000 to a city with 10,000 would raise productivity by over 70%. The third column shows the strength of urbanisation economies by considering how productivity is affected by changes to an urban diversity index (calculated using the shares of different industries in each of the urban areas). As the table shows, for most industries, there is no evidence of significant urbanisation economies. However, for high tech industries, a one standard deviation increase in the diversity index increases productivity by almost 60 percent.

Table 4: Magnitude of Productivity effects for the Republic of Korea

Industry	Localisation economies	Urbanisation economies
Traditional (food, textiles, apparel, wood and paper products and furniture)	0.021*	N/A
Heavy (basic and fabricated metals, chemicals and plastics)	0.082*	N/A
Transport equipment	0.096*	N/A
Machinery and electrical machinery	0.053*	N/A
High tech (computers, communication, television, radio and scientific instruments)	0.056*	59.9*

*significant at 5%. See text for interpretation of coefficients.

Source: Henderson, Lee and Lee (2001)

These findings on urbanisation and localisation appear to be reflected in the pattern of economic activity across cities in developing countries. Activities that are subject to diversification economies tend to be found in the largest cities. Those that are subject to large localisation economies in a few medium size cities, while those subject to smaller localisation economies are less concentrated across a number of small size cities. Thus Lall et al (2004) report that: “We find evidence of high spatial concentration for the Leather and Metals sectors, and moderate concentration in Food Products, Textiles, Mechanical Machinery and Computing and Electronics. Firms in the Paper Products and Chemicals sectors do not exhibit patterns of spatial concentration.”

Evidence on localization is also provided by numerous cases studies of spatial clusters of firms.¹² This literature makes it clear that the clustering of firms to benefit from some sort of agglomeration externalities is a widespread occurrence in developing countries. There are often strong market based input-output linkages between these firms and some evidence that the non-market exchange of goods, information and people also occurs. Many of these case studies also stress the importance of common cultural and social background in generating particular norms

¹² Examples include: the Sinos Valley, Brazil (shoes); the Gamarra region of Lima, Peru (clothing); Guadalajara and Leon, Mexico (shoes); Eastlands, Kenya (garments); Kamukunji, Kenya (metal products); Ziwani, Kenya (vehicle repair); Lake Victoria (processed fish); Suame, Ghana (vehicle repair and metal work); Western Cape, South Africa (clothing); Tiruppur and Ludhiana, India (knitwear); Agra, India (knitwear) and Sialkot, Pakistan (surgical

of behaviour and of local private and public institutions that might reinforce those norms. However, identifying the true benefits to firms located in these clusters will involve much more analysis, focusing in particular on the identification of a set of firms outside the cluster that provide a suitable set of controls for comparison. (See Visser, 1999 for an attempt to move in that direction). For this reason, we do not consider this literature further here, but instead refer the interested reader to two special issues of *World Development* (Humphrey, 1995 and Schmitz and Nadvi, 1999) for further discussion.

Productivity effects in the informal sector

The evidence on agglomeration effects reported in Table 3 comes exclusively from data concerning the “formal” sector. How should the existence of a large pool of unemployment/underemployed workers (a la Harris-Todaro) change our thinking on agglomeration economies? It might be expected to reduce the benefits of city scale, yet there is no evidence that this is the case (see e.g. Au and Henderson (2004) results for China). There are two possible explanations. One, the existence of an informal sector drives up urban costs and crowds out the formal sector, but not fast enough to offset the positive productivity effects of increasing city size. The second possibility is that the informal sector also contributes to agglomeration economies. This possibility is ruled out by assumption in the Harris-Todaro model but at least some of the evidence emerging from developed countries suggests that this assumption is invalid. In particular, there is evidence that points to the existence of networks of small firms that benefit considerably from the productivity effects of the concentration of employment. In developing countries, authors such as Mukherjee (1990) emphasise the vitality of the informal sector. The informal sector often plays an important and visible role in the case study literature on clusters in developing countries. For example, Chari (2004) in his work on the knitwear cluster in Tiruppur paints a vivid picture of the journey taken by cotton thread through the various milling operations, dyeing firms and fabrication units. Formal and informal sectors play their part along the way as independently owned bullock carts shuttle yarn and knitted cloth between knitwear companies and fabrication units. To assume that no agglomeration externalities exist for Tiruppur’s informal sector and for informal firms more generally is surely inappropriate.

instruments). Giuliani, Pietrobelli, and Rabellotti (2004) list a further forty examples from Latin America covering a diverse range of sectors including footwear, furniture, tiles, automotive industries and software.

Unfortunately, our reading of the econometric studies is that more formal evidence on this issue is simply unavailable. Clearly, this is an important area for future work.

4.2 Costs.

There is surprisingly little systematic evidence about the effect of city size on living costs in developing countries. Evidence from Latin America (see Thomas, 1980 and Henderson, 1988) finds that costs of urban living increase with city size. Moving from a small urban area to a large urban area at least doubles the cost of living. Richardson (1987) finds that the per family marginal investment cost is three times larger in urban than rural areas for four developing countries, Bangladesh, Egypt, Indonesia and Pakistan. The costs are even higher in the primate city. Despite the relative paucity of evidence, it seems uncontroversial to assert that costs do rise with city size. We could do with much more systematic evidence on exactly how these costs vary.

What about the non-economic costs of living in cities? Environmental issues would appear to be to the fore here. The problem is, that while it is clear that these problems are worse in cities than in rural areas, it is not clear how these externalities change with city size. Glaeser (1998) presents evidence for developed countries that suggest that levels of sulphur dioxide and ozone are not related to city size but particulate concentration increases with city size. Shukla and Parikh (1992) suggest that sulphur dioxide has a slight tendency to increase with city size for developing countries. They do not consider the impact on ozone levels. Their results for particulates in less developed countries suggest an inverted-U, first rising then falling with city size. It is very likely that this reflects the absence of industry or income controls (larger cities are richer and do less manufacturing, more services).

4.3 Real incomes

Real incomes give the difference between productivity effects and urban costs, and can be observed directly. In two recent papers, Au and Henderson (2004a, 2004b) adopt this approach to the issue of the how costs and benefits change with city size. The idea is to look directly at the impact of city size on real incomes (the difference between wages and cost of living). The problem with this approach is that most institutional arrangements tend to leave the equilibrium city size somewhere to the right of the optimal point (i.e. on the down side of the *ASB* curve, as

discussed in section 2.1). If that is the case, then looking at real wages as a function of city size should give us a negative relationship. In order to identify the inverted-U shape it is necessary to have in the sample some cities that potentially are smaller than optimal. However this would need a set of institutions that prevent cities from expanding beyond this size. Au and Henderson (op. cit.) suggest that the recent migration restrictions in China provide just such a set of institutions. They thus use data from China to directly study the relationship between real wages and city size.

The results from these two papers, although preliminary, provide several insights which have important policy implications. First, there is an inverted-U between real income and city size. The optimal point on this inverted-U depends on what kind of activity is located in the city (which is not surprising given that, as discussed above, different economic activities have different production effects but we would not expect the relationship between costs and city scale to differ substantially across activities). Second, the exact level of optimal city size is very imprecisely determined. This suggests that policy based on any notion of optimal city size is likely to face insurmountable difficulties in deciding what that optimal city size actually is. Third, at least from an economic viewpoint, it is much more costly to be under-sized than over-sized. This point is so important that it is worth quoting Henderson at length: “For [a city with a manufacturing to service ratio of 1], from a peak size of 1.17m if one subtracts 880,000 people, real output per worker falls by 16%; but, if one adds 880,000, it only falls by 6%. To get a 16% loss on the right side of the peak, one would need to increase employment size by 150% of its size at the peak. Real output per worker has a long flat portion near the peak, and real output per worker drops very slowly past the peak.” (Au and Henderson, 2004b, p. 27). Finally, the very flat peak has implications for city sizes when workers are allowed to be mobile. In particular, cities with slightly better amenities or market access could end up very large compared to cities with slightly worse amenities or market position.

5. Evidence: urban structures, primacy and performance.

The evidence that we present above has been based on the costs and benefits of city scale. In addition to this, it is possible to look at the overall structure of cities in a country. What

empirical regularities are there, and what factors shape this structure? What are the implications of variations in structure, eg for growth and job creation?

5.1 *Urban primacy*

There is a substantial literature pointing to the dominance of the ‘primate’ city in developing countries. Such cities are in some cases more dominant than were the largest cities in developed countries at a similar stage of development. Their dominance tends to increase then decrease somewhat during the course of development (Williamson 1965, Shishido and Wheaton 1982).

The determinants of primacy have been studied by a number of authors. Ades and Glaeser (1995) present evidence that rent-seeking plays an important part in determining the degree of urban primacy. They consider a cross section of 85 countries and examine a variety of variables that determine the degree of urban primacy. They find that political factors play a more important role than economic factors. Urban primacy is associated with high levels of political instability (as evidenced by coups, revolutions and strikes) and with centralised political systems (reflected in voting procedures and other political rights). Thus, everything else equal, primate cities in unstable dictatorships are 50% larger than those in stable democracies. Economic factors also play a role. High tariff barriers encourage urban primacy, possibly because they encourage domestic firms to focus on large domestic markets rather than foreign markets (see Krugman and Livas, 1996). In contrast, intercity transport networks lead to less urban primacy, presumably because they allow firms to serve the large domestic market in the primate city from more remote locations. These findings are reinforced by those of Henderson (2002). He finds that, everything else equal, primate cities that are *capital* cities are 25 percent larger than if those primate cities that are not the centre of political power. A second measure of the centralisation of political power (the share of central government in total non-defence government consumption) also has strong effects on primate city size. Henderson (2002) does not consider external trade tariffs, but he does try a range of infrastructure measures. Like Ades and Glaeser, he finds that dense road networks (the ratio of length of roads to national land) decrease primacy although the effect is insignificant. The number of telephone lines per-capita also has a negative but insignificant effect. One measure of infrastructure does have a negative *significant* impact on the size of the largest city: the density of waterways. Taken together, the two papers provide some tentative evidence that investment in infrastructure could play a

considerable part in reducing urban primacy. The most robust finding is that political decentralisation goes hand in hand with urban decentralisation, presumably because of the reduced incentive to locate in the primate city in order to engage in rent-seeking behaviour.

Two key question issues remain unresolved. First, neither of these empirical papers offer a serious alternative hypothesis for the primacy of a particular city in the absence of rent seeking. Such alternatives are available from the theoretical literature; for example, Puga (1996) uses a simple new economic geography to explain urban primacy as a function of transport costs and the elasticity of labour supply from the agricultural to the urban production sectors. Further empirical work testing such alternative hypotheses against those based on rent-seeking behaviour are needed before we are able to assess the extent to which primacy is explained by these competing hypothesis. The second unresolved issue relates to the impact of this urban primacy on development and growth.

5.2 *Urban primacy and economic growth*

Henderson (2003) provides some preliminary evidence on this second issue. He shows, as in Williamson (1965), that excessive urban primacy has a significant negative impact on economic growth. The relationship between primacy and growth is inverse-U shaped, so that negative effects arise when a primate city is above its optimal size. However, as we discussed in Section 4.3, optimal city sizes are very hard to identify. Indeed, in terms of current economic performance, the results from Au and Henderson (2004) suggest that even being substantially above optimal size only has a fairly small effect on productivity. Thus, it is unclear where these findings of a significant negative impact on growth are coming from unless primate cities really are very large indeed relative to their optimal size. This is possible, but there is also another possibility which down-plays the *causal* nature of the link between primacy and economic growth. Recall, that in section 5.1 we suggested that the evidence points to a rent seeking as an important determinant of urban primacy. *If* urban primacy is largely caused by rent-seeking behaviour, then it may be that *both* slow growth and urban primacy are the negative consequences of that rent-seeking behaviour. Determining whether the relationship between urban primacy and poor economic growth is a matter of correlation or causation is very important. If the underlying problem driving both primacy and slow growth is rent seeking behaviour then a policy targeted at reducing the extent of urban-primacy is unlikely to have any

substantial impact on growth. Such a policy could even have a negative impact if it involves relocating productive economic sectors away from the primate city, because these productive activities may no longer be able to benefit from the urbanisation economies that they enjoyed in the large urban environment. Au and Henderson's (2004) results on the high costs of undersized cities urges particular caution when considering policies to force de-concentration of productive activities from primate cities. This said, there are circumstances where the right policy may well be aimed at de-concentration an issue which we discuss further in Section 6.2 below.

6. Some policy implications

We have argued that cities contain many market failures, and that there are many reasons to expect that unregulated outcomes will be socially inefficient. The failures have two sorts of effects. One is that – given a size distribution of cities – they impede the efficiency with which cities operate. The other is that they distort the size structure of cities sometimes, possibly contributing to excess primacy. Policy responses can also be thought of at two different levels. One is to identify the market failures and target policies to overcome them – the classical instruments being taxes, subsidies, or regulation. The other is to identify the broader institutional failures that prevent private agents (for example firms and city governments) from acting to internalise at least some of the externalities. We discuss some of the difficulties and opportunities for each of these approaches in turn.

6.1 Identifying market failures

As we have seen, there are many types of externalities, both positive and negative, that may be important for policy decisions in the urban context. If these are to be targeted by policy actions, they must first be accurately identified and quantified. While a large body of literature now finds evidence of positive externalities in the relationship between city size and productivity, a number of issues remain. One is addressing adequately the heterogeneity of both workers and firms. If intrinsically productive workers or firms tend to choose to locate in urban areas, then econometric estimates would tend to *overestimate* the size of productivity externalities. Having more detailed data on firms allows us to control for some of these differences, and thus to get a more accurate estimate of productivity effects. Analysis for France by Combes et al (2004)

shows that this selective sorting substantially reduces, but does not eliminate, estimates of the magnitude of productivity externalities.

We know relatively little about production externalities in the service sector. This is clearly important in the developing country context given the role wholesale and retail play in the economies of developing country cities. The approaches adopted for manufacturing are applicable to services although the difficulties with measuring productivity for service sectors may point to a focus on employment rather than productivity. Evidence from developed countries suggests that urbanisation economies may be more important than localisation economies for service sector activity (Henderson et al, 1995). We know relatively little about dynamic externalities, i.e. the impact of urbanisation or localisation in the long run. Examples abound, but we lack systematic evidence. A starting point would be to examine firm life cycles along the lines followed by Duranton and Puga (2001). If the incubator role of cities in terms of cost discovery is important we would expect to see large numbers of start ups in diverse urban environments, with more production moving out to lower cost locations once production becomes more standardised. This is an area where more systematic case studies would also help. More systematic case studies may also be needed to examine the production externalities that occur in the informal sector (see also the discussion in box 2). There is a huge literature on the economics of the informal sector, but we have found little evidence of any consideration of the existence of production externalities. The underlying assumption appears to be that production externalities in the informal sector are zero or negative. This is a crucial assumption and has significant implications for whether, say, restricting urban-rural migration and focusing on rural-urban balance is a sensible policy approach for dealing with urban primacy.

On the other side of the equation are the negative externalities of congestion and pollution. These may well be easier to observe and measure than the positive production externalities discussed above – delays can be counted and pollutants measured and evaluated. This may prove fortunate as well known examples, such as Mexico City, suggest that these costs may be particularly acute in some cities.

What does this suggest for the design of targeted policy? Fairly clear recommendations can be made in some cases. For example, if we have evidence that a particular heavily polluting industry tends to benefit from localisation rather than urbanisation economies then it should be

located in a large diverse city. If urban land markets do not work, then reform to make them function is desirable.

In other cases it is much more difficult. Clearly, it is neither feasible nor desirable to seek to identify and to subsidise every possible source of positive externality in production. The theory of the second best warns us to the dangers of piecemeal policy – the possibility that when there are multiple distortions correcting some, but not all of them, does not necessarily raise welfare. And notions of targeting city size as a whole are fraught with danger. At least conceptually, it should be possible to identify an optimal (or efficient) city size. The available evidence suggests that this is *extremely unlikely* to provide a good policy target in practice. This reflects that fact that there are very large margins of errors associated with attempting to identify optimal city size. Theoretical and empirical evidence does suggest that, in most institutional contexts, moving towards the unknown optimal city size would involve attempts to constrain the size of the larger cities or facilitate the growth of new cities. The key problem with this as a policy objective is that available evidence suggests that, at least from an economic perspective, being oversized is much less costly than being undersized. Work on China suggests that getting it wrong and restricting city size too much can lead to fairly large economic losses. Given the uncertainty involved, limiting city size on the basis that the optimal city size may be smaller would appear to be a highly risky policy objective.

6.2 Institutions to promote efficiency

Standing back from the detail, one of the main issues that we have identified is the disadvantage that a firm might face in locating outside an existing centre of activity, and the consequent bias towards primacy that this creates. We pointed to different factors creating this disadvantage. One was the existence of institutional structures that encourage the rent-seeking and thereby support large primate cities. Another was uncertainty about the future growth prospects of secondary cities – an ‘isolation paradox’ or coordination failure deterring firms from relocating. A third arises in public finance, as developing cities typically require large public investments, yet lack both the tax base and the ability to borrow to finance such investments. We also saw how poor inter-city infrastructure and restrictive international trade policy tend to promote primacy.

These are issues that can, potentially, be addressed by institutional reform. Deregulation and measures against corruption can reduce the attractiveness of the primate city as a source of rents. City governments can be equipped with powers and incentives to plan and finance development in cooperation with the private sectors; while devolving borrowing powers clearly has risks, attention needs to be paid to designing mechanisms that can fund the required investments. National infrastructure policy can be shaped with a view to facilitating deconcentration.

An interesting example of the effectiveness of alternative policies is provided by the case of Korea (see Henderson, Lee and Lee 2001) in which activity was deconcentrated away from the Seoul region. Initial heavy concentration in the Seoul region had been encouraged by state regulation and by poor infrastructure investments in outlying districts. In the early and mid 1970s policies were introduced that included financial incentives for relocating firms, direct relocation orders and public construction of infrastructure parks. They had limited success, only encouraging decentralisation from central Seoul to areas within 45 minutes drive from the city. In the early 1980s policy turned to wider deregulation, reducing the need to be close to Seoul to deal with government. In addition, there was a massive investment drive to improve infrastructure and communications to a designated set of primary and secondary cities. This combination was successful in reducing the share of manufacturing employment in the 3 largest urban areas from 44% in 1983 to 28% in 1993.

In conclusion then, policy design needs to be thought through at two levels. One is identifying the market failures and, in some cases, using targeted micro-economic policy to address them. The other is recognising that these market failures may create systematic distortions in urban structure. They can be mitigated by the design of infrastructure policy and by rather broad policies of institutional reform and decentralisation.

References

- Ades, A.F. and E.L. Glaeser (1995), 'Trade and Circuses: Explaining Urban Giants,' *Quarterly Journal of Economics*, 110, 195-227.
- Arnott, R.J. and M. Gersovitz, (1986) 'Social Welfare Underpinnings of Urban Bias and Unemployment' *Economic Journal* 96, 413-424
- Au, Chun-Chung and Vernon Henderson, (2004a) How Migration Restrictions Limit Agglomeration and Productivity in China, Processed, Brown University.
- Au, Chun-Chung and Vernon Henderson, (2004b) 'Estimating Net Urban Agglomeration Economies With An Application to China', Processed, Brown University.
- Bairoch, Paul (1988) *Cities and Economic Development: From the Dawn of History to the Present*. Translated from the French by Christopher Braider. University of Chicago Press, Chicago.
- Becker, C.M. and A.R. Morrison, (1999) 'Urbanisation in transforming economies' in Cheshire, P. and E.S. Mills (eds) '*Handbook of regional and urban economics*' vol 3, North Holland.
- Black, D. and J. V. Henderson (1999b), 'A Theory of Urban Growth', *Journal of Political Economy*, 107, 252-284.
- Brueckner, J. K (1990), Analyzing Third World Urbanization: A Model with Empirical Evidence, *Economic Development & Cultural Change*, 38(3) 587-
- Brueckner, J. K and Y. Zenou (1999), 'Harris-Todaro models with a land market' *Regional Science and Urban Economics*, 29, 317-339.
- Chari, S. (2004) *Fraternal Capital: Peasant-Workers, Self-Made Man and Globalization in Provincial India*, Permanent Black, Delhi.
- Chen, Y. (1996) Impact of regional factors on productivity in China, *Journal of Regional Science*, 36: 417-436
- Combes, P. P., G. Duranton, and L. Gobillon (2004) Spatial wage disparities: Sorting matters. Processed, London School of Economics
- Cronon, W. (1992), *Nature's Metropolis: Chicago and the Great West*, W. W. Norton & Co., New York.
- Duranton, Gilles and Diego Puga (2001) Nursery Cities: Urban diversity, process innovation, and the life cycle of products. *American Economic Review* 91(5): 1454-1477.

- Duranton, G. and D. Puga (2004) Micro Foundations of Urban Agglomeration Economies, in V. Henderson and J. Thisse (eds.) *Handbook of Regional and Urban Economics*, Vol 4., North Holland.
- Eaton, J. and Z. Eckstein (1997), 'Cities and Growth, Theory and Evidence from France and Japan,' *Regional Science and Labor Economics*, 27, 443-474.
- Eberts, R.W. and D.P. McMillen 'Agglomeration economies and public infrastructure', in Cheshire, P. and E.S. Mills (eds) '*Handbook of regional and urban economics*' vol 3, North Holland.
- Fujita, M., P. Krugman and A. Venables (1999) *The Spatial Economy: Cities, Regions and International Trade*, MIT press.
- Glaeser, E. (1998) Are cities dying? *Journal of Economic Perspectives*, 12(2) pp. 139-160.
- Hausmann, R., and D. Rodrik. (200). "Economic Development as Self-Discovery." Working Paper no. 8952. National Bureau of Economic Research, Cambridge, MA.
- Henderson, V. (1974) "The Sizes and Types of Cities," *American Economic Review*, 64, 640-656.
- Henderson, V. (1988) *Urban Development: Theory, Fact and Illusion*. Oxford University Press: New York.
- Henderson, V. (2002) Urbanisation in developing countries, *The World Bank Research Observer*, 17(1) 89-112.
- Henderson, V. (2003) The urbanisation process and economic growth: The so-what question *Journal of Economic Growth*, 8 47-71.
- Henderson, V. (2004) Urbanisation and growth" forthcoming, *Handbook of Economic Growth*, eds P. Aghion and S. Durlauf, North-Holland.
- Henderson, V. and J. Thisse (eds.) (2004) *Handbook of Regional and Urban Economics*, Vol 4, North Holland.
- Henderson, V. and Kuncoro (1996) Industrial Centralization in Indonesia, *World Bank Economic Review*, 10:513-540.
- Henderson, V, Kuncoro, and M. Turner (1995) Industrial development in cities, *Journal of Political Economy*, 103, 1067-1090.
- Henderson, V. Lee, and Lee (2001) Scale externalities in Korea, *Journal of Urban Economics*, 49: 479-504.

- Henderson, V and A. J. Venables (2004) 'The dynamics of city formation: finance and governance', CEPR dp no 4638.
- Humphrey, J. (1995) Introduction, *World Development*, 23(1) pp. 1-7.
- Kessides, C. (2004), 'The contributions of urban development to economic growth and poverty reduction – with special focus on Sub-Saharan Africa', World Bank, Washington.
- Krugman, P. R. and R. Livas, (1996) Trade Policy and the Third World Metropolis, *Journal of Development Economics*, Vol 49, pp. 137-150.
- Lall, Somik, Jun Koo and Sanjoy Chakravorty, (2003) Diversity Matters: The Economic Geography of Industry Location in India, *World Bank Policy Research Working Paper Series #3072*.
- Lall, S. Shalizi, Z and Deichmann, U (2004) Agglomeration economies and productivity in Indian industry, *Journal of Development Economics*, 73 pp 643-673.
- Lee and Zang (1998) Urbanization and regional productivity in Korean manufacturing *Urban Studies*, 35, 2085-2099.
- Lipton, M (1976) 'Why poor people stay poor; urban bias in world development', Harvard University Press, Cambridge MA
- Lipton, M. (1993) 'Urban bias; of consequences, classes and causality', *Journal of Development Studies*, 29, 229-258
- Mitra (2000) Total factor productivity growth and urbanization economies: a case of Indian industries, *Review of Urban and Regional Development Studies*, 12, 97-108.
- Mukherjee (1990) Conditions of labor in the small scale and unorganised sectors in Calcutta and its neighborhood, WEP 2-43/WP.36 International Labour Office, Geneva.
- Nadvi, K. (1999) The Cutting Edge: Collective Efficiency and International Competitiveness in Pakistan, *Oxford Development Studies*, 27 (1), 81-107.
- Nash, M. (ed) (1977) 'Essays on economic development and cultural change; in honor of Bert F. Hoselitz'; University of Chicago Press.
- Richardson, H (1987), The costs of urbanisation: A four country comparison" *Economic Development and Cultural Change*, 33: 561-80.
- Rosen, K. and M. Resnick (1978), 'The Size Distribution of Cities, the Pareto Law, and the Primate City,' mimeo, Princeton University.
- Rodrik, Dani (2004) *Industrial policy for the twenty-first century*, CEPR Discussion Paper no. 4767.

- Rosenthal, Stuart and William Strange (2004) 'Evidence on the nature and sources of agglomeration economies' in J. V. Henderson and J. Thisse (eds.) *Handbook of Regional and Urban Economics*, Vol 4, North Holland.
- Schmitz, H. and K. Nadvi (1999) Clustering and industrialization: Introduction, *World Development*, 27 (9) 1503-1514.
- Shishido, H. and W. Wheaton (1981), 'Urban Concentration, Agglomeration Economies, and the Level of Economic Development,' *Economic Development and Cultural Change*, 30, 17-30.
- Shukla (1996) *Urbanisation and Economic Growth*, Delhi: Oxford University Press India (Described in Handbook, volume 3 p.1760)
- Shukla, V. and Parikh, K.,(1992) The environmental consequences of urban growth: cross-national perspectives on economic development, air pollution, and city size. *Urban Geography*., 12(5): 422-449.
- Sveikauskas, L. (1978), 'The Productivity of Cities', *Quarterly Journal of Economics*, 80, 393-413.
- Tacoli, C (2004), *The Role of Small and Intermediate Urban Centres and Market Towns and the Value of Regional Approaches to Rural Poverty Reduction Policy*. Paper prepared for the Prepared for the OECD DAC POVNET Agriculture and Pro-Poor Growth Task Team
- Thomas, V (1980) 'Spatial differences in the cost of living', *Journal of Urban Economics*, 8: 108-22.
- Venables, A.J. (2005), 'Spatial disparities in developing countries: cities, regions, and international trade' *Journal of Economic Geography* 5: 3-21
- Visser (1999) A Comparison of Clustered and Dispersed Firms in the Small-Scale Clothing Industry of Lima, *World Development*, 27 (9), 1553-1570.
- Williamson, J. (1965) Regional inequality and the process of national development, *Economic Development and Cultural Change* 13, 3-45.
- World Bank (1999), *Entering the 21st Century, World Development Report 1999/2000*, Oxford University Press.
- World Bank (2005) *Doing Business in 2005: Removing Obstacles to Growth*. A co-publication of the World Bank, the International Finance Corporation and Oxford University Press