SESSION 2

LIMITS TO POPULATION GROWTH?

6 Planning sustainable cities in the 21st century

Richard Arnott[1]
University of California, Riverside

What problems related to metropolitan spatial growth are likely to be encountered in Australian cities in the 21st century? Anticipating these problems, how should the federal, state and local governments plan to deal with them? Will metropolitan growth be consistent with 'sustainability' (which I use imprecisely as 'steady improvement in the quality of life')?

I am not comfortable with 'Big Think'. I decided instead to write up a series of 'Little Thinks'. This paper will therefore lack the coherence of Big Think, but will hopefully contain a few ideas that are interesting and practically useful.

Let me start with a statement of 'where I'm coming from'. Intellectually, I am very much a child of my generation. In the 1960s, there was an explosion of interest in urban problems that grew out of the idealism and radicalism of the time: Lyndon Johnson's Great Society program, Michael Harrington's *The Other America* (1962), the civil rights movement, Jonathan Kozol's *Death at an Early Age* (1967), and the urban race riots. The low point for US central cities was the 1970s and 1980s, but in the 1960s the rot had clearly set in. The white flight to the suburbs, hastened by blockbusting, was almost over. Blacks were excluded from most suburbs by racial discrimination. Many of the inner suburbs had tipped from being 90 per cent white to 90 per cent black. Large segments of central cities became crime-ridden ghettoes, and the situation was made worse by the sharp fall in downtown property values, which led to the erosion of the property tax base and a decline in the quality of public services (Mieszkowski and Mills 1993).

My undergraduate degree was in 'urban studies'. MIT had no undergraduate program in urban studies at the time, so I put together my own program with an

1 I would like to thank John Daley and Kevin O'Connor for their valuable comments as discussants of the paper at the Roundtable.

amalgam of courses in urban planning (then only a graduate program), architecture, civil engineering and economics. When I started graduate studies in the early 1970s, the general equilibrium revolution in applied microeconomic theory was in full swing. It coloured the perception of my generation of economics graduate students in two distinct ways. First, it caused us to look at economic problems from the perspective of the entire economy rather than from the perspective of single-market partial equilibrium analysis. Second, its foundation is axiomatic general equilibrium theory (Debreu 1959), which assumes all markets to be perfectly competitive, and whose central result is the First Theorem of Welfare Economics. Even though we were made aware of the restrictiveness of the theory's assumptions, through the theory of market failure (Bator 1958; Salanié 2000) and the theory of the second best (Salanié 2003), we nonetheless came away with a deep appreciation for the virtues and even the beauty of organising the economy through markets.

Urban economics grew out of the mixing of these two developments — reformist concern for the plight of the American central city, and the general equilibrium revolution in applied microeconomic theory. And their mixture accounts for the field's peculiar combination of left-wing concerns and qualified faith in the efficacy of markets, and for its benevolent planner perspective and lack of emphasis on political constraints.

This paper addresses facets of the opening questions from the perspective of urban economics and from the perspective of developed countries. Most of the urban drama in the 21st century will take place in developing countries, where urban populations are projected to increase from 2 billion in 2000 to 4 billion in 2030 and 5.5 billion in 2050 (Angel et al. 2010). In developed countries, in contrast, the urban population is forecast to stabilise at around 1 billion. In being land rich, Australia, Canada and the United States constitute a distinct subset of the developed countries. Since there is considerable uncertainty about how much immigration each country will choose to permit, metropolitan planning will need to accommodate alternative population scenarios.

I apologise in advance for the provincialism of my essay. It draws little on the Australian experience and perhaps too much on the Los Angeles experience.

6.1 Trends in urban spatial structure

Over the past decade, satellite imaging has revolutionised the measurement of urban spatial structure (Angel et al. 2010; Burchfield et al. 2006). We can now estimate coverage and floor-area ratios at a high degree of resolution and with reasonable accuracy across an entire metropolitan area.

In almost all cities, population and average income are increasing, and the money price of driving is decreasing. These factors together have generated strong empirical regularities with respect to trends in urban spatial structure.

1. Between 1990 and 2000, the mean population density of a global sample of 88 large 'cities' (metropolitan areas) declined at an average annual rate of 2 per cent (Angel et al. 2010). For a sample of 20 US cities between 1910 and 2000, average annual density decline was 2 per cent, too.
2. For a global sample of 119 large metropolitan areas in 2000, the elasticity of average population density with respect to population is 0.20, of average population density with respect to per capita income is -0.40, and of average population density with respect to the gas price is -0.13 (Angel et al. 2011, Table V, Model 3).
3. The population density gradient (Clark 1951 — the proportional rate at which population density falls with respect to distance from the metropolitan centre) has been falling consistently in virtually all metropolitan areas since data were available (widely documented, and reported in Anas, Arnott and Small 1996).
4. In US cities, suburbanisation started in the mid-19th century, with the advent of the railway suburb. The suburbanisation of the population continued steadily until World War II, and then accelerated in the immediate postwar period. The suburbanisation of jobs lagged the suburbanisation of the population. By 1990, in the 100 largest metropolitan areas in the United States, the average household lived 8 miles from the metropolitan centre (Glaeser and Kahn 2001), while the average job was located 7 miles away. Of course, these distances increase with the metropolitan area population. The important regularity is that jobs are now almost as decentralised as residences.
5. Metropolitan areas are becoming increasingly polycentric (Anas, Arnott and Small 1998; Gordon and Richardson, 1986). An employment subcentre is defined as a set of contiguous zones (census tracts, traffic analysis zones etc.), each having an employment density exceeding d per unit area and having combined employment exceeding D. Within a typical metropolitan area, as time proceeds employment subcentres become more numerous, and the ratio of aggregate employment in employment subcentres outside the traditional central city centre to that in the traditional city centre increases. One may employ alternative definitions of subcentres, and the same applies.
6. One can measure employment dispersion in terms of the ratio of employment outside subcentres to total employment. By this measure, employment is becoming increasingly dispersed (Gordon and Richardson 1996).

7. Over the past 50 years, US metropolitan areas have been steadily less segregated in terms of race and religion, and steadily more segregated in terms of income (Madden 2003).

6.2 Empirical regularities with respect to travel

Trip tours and trip chaining are becoming increasingly important. When looking at the data, one must therefore pay careful attention to how a trip and the nature of a trip are classified. An increasingly small proportion of travel is on the journey to and from work, whether directly or as part of a trip chain.

The era of massive freeway building ended around 1970. Since then, the population of US metropolitan areas has increased steadily, as has mean income. With the monocentric city model in mind, for auto commuters one would expect average commuting distance to have increased, average travel speed to have decreased, and average commuting time to have increased. Paradoxically, however, while average commuting distance has indeed increased, average travel speed has increased, so that average commuting time has increased only slightly. Making the paradox even more paradoxical, the average time lost due to congestion has been increasing steadily (Gordon and Richardson 1989; Texas Transportation Institute 2010)

The paradoxes have not been completely resolved, but three elements seem to be particularly important. First, since jobs have suburbanised, an increasingly large proportion of commutes are from suburb to suburb. Second, travel speeds in uncongested to moderately congested traffic have increased significantly due to improvements in automobile safety and design. Third, the length of the rush hour has increased.

One can imagine several scenarios in which a metropolitan area's population and average income grow at a constant rate, and yet commuting times remain constant. In the first, which I present to my undergraduate class, one can imagine a situation in which the city remains monocentric and commuters will not tolerate a commute of more than 30 minutes. Combining this simple rule of thumb with technological improvements in transportation (the horse-drawn streetcar, the electric streetcar, the subway, the bus, and the car) actually goes a long way to explaining the evolution of urban spatial structure in US cities over the past two hundred years (Warner 1962). In the second, the city remains monocentric and investment in transportation infrastructure and technological improvements in transportation are such that mean commuting time remains constant. In the third, the city remains monocentric and the length of the rush hour increases. In the fourth, the metropolitan area expands outwards with no improvements to the existing road system but with continual

additions to it at the urban periphery (Gordon and Richardson 1989). Employment subcentres emerge at the urban periphery at such a rate as to keep mean commute times unchanged. In this sense, there are several sustainable ways in which a metropolitan area can keep growing in population and getting richer.

6.3 Go with the flow: rechannelling, not opposing, market forces

Economists and planners often disagree. Economists have qualified faith in the operation of markets and favour pricing solutions. Planners, especially those in the British tradition, are sceptical of the market, find many aspects of market outcomes unlovely (in the two senses of unjust and unaesthetic) and have traditionally favoured regulation. Over the course of my career, there has been a partial convergence. When I started teaching, planning professors at Queen's University urged their students not to take my urban economics course on the grounds that I would corrupt them. That would no longer happen. The ideological divide between Marxists and non-Marxists is no longer at centre stage. And the historical experience of the past 75 years has persuaded many of those who were attracted by socialist ideals that market-based economies, warts and all, tend to work better than planned economies. At the same time, the recent financial crisis has caused economists to be less arrogant and to acknowledge that economic liberalism can be carried too far.

There seems to be a broad consensus that economic policy generally and metropolitan planning in particular should *harness* or *rechannel* market forces rather than oppose them, though there remains considerable disagreement about how principle should be translated into practice. In the context of economic development, the Washington Consensus of the 1980s and 1990s has largely broken down, and the World Bank is now seeking a middle way, as evidenced by the work undertaken under the aegis of the Commission on Growth and Development (which includes a volume on urbanisation and growth).

The virtues of markets are manifold. Markets deliver goods to those who value them the most. Markets solve a huge coordination problem with low transaction costs. Markets efficiently aggregate information (at least under ideal conditions). When prices are right, markets create incentives for individuals to act in the social interest (Adam Smith's 'Invisible Hand' — Smith 2009). Markets provide the incentives for individuals to work hard, to work on those things that are socially valued, and to undertake the invention and innovation that underlies economic growth. Markets are compatible with consumer sovereignty, individual freedom, civil liberties and democratic institutions.

But markets also fail in small ways and in large. Even when markets operate efficiently, their outcomes may be highly unjust and inequitable, and distasteful in other ways. Markets are also subject to the classic market failures — public goods, externalities and increasing returns to scale (Bator 1958) — as well as the more modern failures associated with asymmetric information and coordination failure (Salanié 2000).

Good economic policy entails redistributing income and equalising opportunity to make society more just, and correcting market failure through taxation, regulation and the creation of new markets.

6.4 In defence of consumer sovereignty

Another traditional bone of contention between economists and planners is respect for consumers' tastes. Microeconomic theory takes consumer tastes, as described by utility functions, as exogenous, and, excepting addictions, respects the economic choices that people make, conditional on their facing prices that reflect true social cost. This may entail a teenager 'wasting' his pocket money on the latest designer running shoes and rap music, or a middle-class family 'squandering its income' on a tasteless and unnecessarily large home, replete with the latest appliances, rather than buying goods that celebrate the beauty of nature or the glory of the human spirit, or giving their money to charities dedicated to social justice. In contrast, many planners, especially those in the British tradition, are willing to impose their view of the good life on others.

Given my tastes, I wish that others did not waste their money. At the same time, given theirs, they likely wish that I did not waste mine. I might feel happier if I could constrain their choices or force them to consume what I think would be best for them, but I would certainly not feel happier if they were to constrain my choices or force me to consume what they think would be best for me. Behind the veil of ignorance, where not only my endowments but also my tastes are revealed after the veil is lifted, I would choose that everyone respect others' tastes rather than have a subset impose their tastes on others.

This line of argument is essentially the same as that which has been used to defend cultural relativism and moral relativism. Taste and cultural relativism lead to stimulating diversity and pluralism. Moral relativism is constrained by the need to have a legal system based on a consistent set of beliefs.

6.5 Getting the prices right

Consider a person who has to decide how many units of a commodity to consume. It is almost a truism that, if he faces the full social cost of each unit and enjoys the full social benefit, he will choose to consume the socially optimal amount. In a perfectly competitive economy, with no externalities and with complete markets, these conditions are satisfied. Under these circumstances, the price of a good is equal to the marginal social cost of its production; thus, when purchasing the good in a competitive market, the individual does indeed face the full social cost. Since he chooses the number of units to purchase such that the benefit he derives from the last unit equals the price, his marginal private benefit equals the price of the good. And since there are no externalities, the marginal social benefit equals his marginal private benefit.

Unfortunately, of course, externalities are pervasive, few markets are perfectly competitive, and many markets are absent. The benevolent planner who aims to improve the performance of a market economy will attempt to correct these market failures through taxation, regulation and the creation of new markets.

There is an ongoing debate within public economic theory concerning how to integrate equity and efficiency. When I started my graduate studies, I was taught that equity can be dealt with through lump-sum redistribution, in which case efficiency can be pursued as an independent goal. It was then pointed out that lump-sum redistribution on the basis of need is impossible, since need is unobservable and, if any proxy of need such as income is employed, individuals will adapt their behaviour in distortionary ways to appear more needy. A set of papers then appeared that argued for distributional considerations to be taken into account in deciding on commodity tax rates, public sector prices and the level of public goods. It was then argued that, under certain not unreasonable conditions, an optimal income tax redistributes income as efficiently as possible, so redistribution by other means is both unnecessary and distortionary.

The current majority view seems to be that commodity tax rates, public sector prices and the level of public goods should be chosen without regard to their distributional effects (Kaplow 2008). It is interesting that William Vickrey, who was very much a champion of the poor, argued for marginal cost pricing, without regard to distributional considerations (Arnott et al 1994).

6.6 Optimal metropolitan spatial structure

To start, consider a large, homogeneous plain that is uniformly endowed with resources and population. To simplify, suppose there is a single, generic consumption good that is produced in factories using the resources and homogeneous labour. The optimal spatial structure maximises the per capita consumption of the good (or, since population density is given, per unit area of land). If there are constant returns to scale in production and if resources are costly to transport to the factory, the optimal spatial structure is a 'backyard economy', in which people are distributed uniformly over space, each operating his own miniature factory using resources from his personal 'market area'.

A 'non-trivial' spatial structure requires either that space be inhomogeneous in terms of transport costs (rivers and oceans) or resource endowments, or that production occurs under increasing returns to scale (these correspond to Cronon's first and second nature; 1991). The earliest cities may have formed on the basis of unequal resource endowments, leading to river trade, combined with increasing returns to scale in loading and unloading, or of increasing returns to scale in grain storage or religious observance. Until the Industrial Revolution, most large cities were port cities that arose on the basis of economies of scale in shipping, with trade deriving from unequal resource endowments. In the early Industrial Revolution, economies of scale internal to the individual plant became of prime importance. With the exception of Rome, whose size was partially attributable to economies of scale in the administration of the Roman Empire and partly to its receipt of tribute, no city up to 1800 had a population exceeding one million (Bairoch 1988). Thus, the general opinion is that neither unequal resource endowments combined with economies of scale in shipping, nor economies of scale internal to the individual plant, can explain the modern city.

Three general factors are responsible for the rise of today's large cities. The first is technological improvements in goods transportation; goods transportation costs have fallen by a factor of 100 in the past two centuries (Glaeser and Kohlhase 2003), and there has been a series of technological improvements in intra-urban people transportation that has permitted the areal expansion of the city (Warner, 1962). The second is technological improvements in building construction. The third is an increase in the complexity of production combined with *external economies of scale.* External economies of scale are economies of scale that are external to the individual plant, but internal to a firm, an industry in a particular city, or the city as a whole (Fujita, Krugman and Venables 1999; Henderson 1974). With external economies of scale, the individual firm's cost per unit of output is independent of how much it produces, but its units cost less the larger the agglomeration in which it is located because of the increased labour specialisation

possible with a larger labour pool, the increased specialisation in intermediate goods possible in a larger city, and collective learning by doing (Marshallian economies of scale). Since the individual firm contributes to the size of the agglomeration and therefore reduces the unit production cost of other firms in the agglomeration, it benefits them. And since the firm is only partially compensated by the other firms, the benefit it confers on them is, partially at least, an uninternalised positive 'agglomeration' externality (Fujita and Thisse 2002). Note that external economies of scale are consistent with each firm being a price taker.

The modern view, as reflected in the literature on the new economic geography (Fujita, Krugman and Venables 1999; Henderson and Thisse 2004), is that the spatial pattern of production, whether at the global, national, regional or metropolitan scale, derives from the interplay between the agglomerative or centripetal force of external economies of scale and the deglomerative or centrifugal force of transport costs.

6.7 Potentially important market failures in metropolitan spatial structure

The centrepiece of the new urban economics is the monocentric city model (Alonso 1964; Brueckner 1987; Mills 1967; Muth 1970; von Thünen 1827). In that model, residential lots surround a point central business district (CBD), and each day each of the fixed number of residents commutes to the CBD to work and shop. In the basic model, production occurs under constant returns to scale and commuting costs are not subject to congestion. The First Theorem of Welfare Economics states, 'Any competitive equilibrium is Pareto optimal.' There is an analogous theorem for the monocentric city model that 'Any monocentric city equilibrium is Pareto optimal' (Mirrlees 1972). This theorem is not implied by the First Theorem since the structure of the economy is different, but its proof is much the same. The theorem is important. First, it indicates that, with constant returns to scale in production at a point CBD and absent congestion externalities in transportation (and other potential sources of externality that are not in the model), in the monocentric model at least, a competitive land market allocates space efficiently. Second, it points to two potentially important sources of market failure in more realistic urban economies — economies of scale in production and uninternalised congestion externalities. Conventional wisdom is that these are indeed the two most important market failures with respect to metropolitan spatial structure.

Auto congestion

When auto congestion is introduced into the monocentric city model, if the congestion externality is internalised, for example through the application of an optimal congestion toll, the economy is efficient. If, however, the congestion externality is not internalised, then transportation is underpriced, and cities become inefficiently sprawling and low density (Kanemoto 1980; Solow 1972).

External economies of scale/agglomeration externalities

Remove auto congestion, add economies of scale in production, and endogenise firm location. Each firm then locates so as to maximise profits, taking as fixed the location of households and other firms. If the economies of scale to firms are internal to firms, the market power problems associated with natural monopoly are encountered, and firm behaviour must be modelled using the tools of game theory. If the economies of scale are external to firms and are localised so that a particular firm's productivity depends on its proximity to other firms, equilibrium exists but its properties are quite different from those of the monocentric city model (Fujita and Ogawa 1982). First, there may be multiple equilibria. Second, for at least some of those equilibria, cities are polycentric. Third, equilibrium is inefficient for the reason noted above — the uninternalised positive agglomeration externality. Furthermore, since agglomeration externalities are by their nature very difficult to measure, unlike the congestion externality, there is no simple policy remedy to internalise them. Remarkably, no-one has undertaken a thorough investigation of how external economies of scale distort metropolitan spatial structure. Imagine plotting the net marginal private benefit to a firm from joining an agglomeration of employment size n compared to operating in isolation. Since the marginal social benefit curve lies above the marginal private benefit curve, too few firms join agglomerations. It is also reasonable to conjecture that, if all subcentres are of the same size and if all firms produce in subcentres, then equilibrium will entail too many subcentres of suboptimal size. But all subcentres are not of the same size, and there may exist multiple equilibria. Adding congestion externalities to the mix further complicates an already very complicated problem.

How should the policymaker respond to this vexing problem? There is no right answer. Like most urban economic theorists, my gut reaction is that the planner should internalise those externalities that he can, but the scientific basis for this prescription is weak.

Migration

Migrants generate both positive and negative externalities. By increasing the number of workers in a city, via external economies of scale a migrant likely increases the productivity of existing workers. And by increasing the population of the city, her presence increases congestion in the city. There may also be fiscal externalities associated with migration. If these externalities are efficiently priced, her private decision to migrate is socially optimal. But, in fact, neither the positive agglomeration externality nor the negative congestion externality is fully internalised.

It is therefore not possible a priori to say whether there is too much or too little rural–urban migration. The older literature on rural–urban migration, which was heavily influenced by the Harris–Todaro model (1970), neglected the agglomeration economy and so argued that rural–urban migration is excessive. Partly for this reason, almost all developing countries have discouraged rural–urban migration. There is now a new view of rural–urban migration, based on the new economic geography, which takes the agglomeration externality into account and sees cities as engines of growth (Duranton 2009). According to this new view, the largest city in a country tends to be too small.

In most developed countries, net rural–urban migration is now small, and the important internal migration occurs between regions and within systems of cities in an urban hierarchy. How the actual distribution of population differs from the optimum then depends on how the size of the uninternalised portion of the net migration externality varies across regions and across cities in the city-size distribution (Papageorgiou and Pines 1999). There is no consensus in the literature on how the equilibrium and optimal city-size distributions differ. The literature therefore suggests that government intervention to alter the city-size distribution is unwarranted.

How rapidly Australia's cities grow over the next decades will depend first and foremost on the level of immigration. Australian policy makers will need to plan their cities' growth under alternative population scenarios.

Pollution

From a conceptual point of view, pollution can be dealt with straightforwardly by charging polluters for the social cost of the pollution they generate. One practical problem is monitoring, but the biggest problem is overcoming the opposition of industrial polluters, particularly during hard times when the threat of job losses is a real one. The Los Angeles metropolitan area has been remarkably successful in

reducing the level of air pollution. If the political will is there, the problems can be managed.

Water

In California, the management of water has been and continues to be a major problem. The problem is a political one. The Central Valley's agriculture consumes a significant portion of the state's water at very low prices, which results in rice being produced in semi-desert conditions. Conceptually, setting the price of water at marginal social cost and following sound cost–benefit practice in managing the water infrastructure system can straightforwardly solve the problem.

Land assembly

Through its exercise of eminent domain, government has an advantage over the market in land assembly. The market failure derives from the inefficiency of the game between the assembler and the property owners (Strange 1995).

6.8 Densification

As a city's population grows and as its residents become richer, the demand for floor area increases. The market deals with this increased demand by a combination of an increase in rent and an increase in the quantity of floor area. The growth in floor area can occur either horizontally, vertically or through infill and add-ons. I live in Riverside County, the population of which almost doubled between 1990 and 2010, from 1.17 million to 2.14 million. The current economic crisis notwithstanding, the population is expected to double again in the next 40 years. Thus, it is natural for me to think of horizontal expansion. But increases in floor area can also be achieved by increases in the floor-area ratios of already developed areas, through redevelopment at higher density, infill and add-ons — a process called *densification.* Vancouver's West End went through the most dramatic densification that I am familiar with, with stately mansions on large lots being replaced by 30-storey residential high-rises. Cities choose whether and how they wish to densify. Vancouver has chosen to permit the densification of some neighbourhoods but not of others. Boston has chosen to allow densification of the city's two downtown cores — the financial district and the Prudential–Copley area — but not of any of the downtown residential areas.

6.9 Zoning issues and other land-use controls

In an undergraduate urban economics examination, I recently posed the question, 'What would be the result of removing zoning?' Almost to a person, the students answered, 'Chaos.' (Evidently, I had not done a very good job in corrupting them). That is not the right answer. Land would go to the 'highest use' — that use which bids the most for it — but because of externalities, not necessarily its best use. Zoning was originally introduced with the intention of separating incompatible land uses.

Most cities in the United States have Euclidean zoning. It is called Euclidean zoning not after the geometer but after *Village of Euclid, Ohio v. Ambler Realty Company*, which established the constitutionality of zoning. Euclidean zoning entails the rigid separation of land uses into residential, commercial and industrial districts (and often separation between subcategories as well). Euclidean zoning was so rapidly adopted because it is easy to implement and does indeed achieve the goal of separating incompatible land uses. But Euclidean zoning is widely criticised today for giving rise to a dull uniformity of residential neighbourhoods, for separating housing and jobs, and for being insufficiently flexible and insufficiently adaptable to changed economic circumstances. Changes in land use are permitted, but through a lengthy process of zoning variances being granted by local zoning boards. Many US cities are now making their zoning policies more flexible, allowing Euclidean II zoning (which some define too specifically as hierarchical zoning), incentive zoning (where variances are granted when a proposed project meets development goals), and smart zoning (which is associated with smart growth, and of which cluster zoning is one type). Neighbourhoods that were dull and uniform at the time of initial construction become more interesting with improvements and more varied land use. A similar argument can also be made for greater flexibility in the design and application of the land use master plan. Flexibility is especially important in metropolitan areas where future population growth is uncertain.

Many suburban towns in the United States use land-use controls to exclude the poor and other undesirable elements. This is referred to as NIMBYism (Not In My Back Yard). When I visited Stanford in 1990, I encountered an extreme example. A wealthy suburb close to Palo Alto zoned out old folks' homes solely on the basis of a hedonic study that found them to reduce property values. The term *exclusionary zoning* refers to the use of zoning with the intention of excluding certain groups from the community. The most familiar method of exclusionary zoning is minimum lot size zoning. In the Boston metropolitan area, some suburban communities where land prices are $500 000 per acre impose minimum two-acre lot sizes. The rich may want to exclude the poor either because, under property taxation, the poor would pay less for the same level of public services (Hamilton 1975; Tiebout 1956), or

simply to keep out the riffraff, who depress property values. Newton uses wetland zoning to exclude affordable housing projects. Exclusionary zoning is unhealthy in separating the rich and poor, and distorts urban spatial structure. It also seems (the phenomenon is well documented empirically but the theory has not been well worked out) to steepen the supply curve for suburban development (Glaeser, Gyourko and Saks 2005; Saiz 2010), not only driving up property prices and rents throughout the metropolitan areas to inefficiently high levels, hurting the poor, but also increasing the amplitude of real estate cycles. In southern California, NIMBYism seems to be less of a problem. There the rich exclude the poor more through the use of gated communities (Blakely and Snyder 1997). Gated communities allow the rich to enjoy their preferred level of public services without subsidising the poor's consumption of them, and keep the riffraff out without impeding suburban expansion.

6.10 Impact fees

An *impact* or *development fee* is a fee that is imposed by a local government on a new or proposed development to cover a portion of the costs associated with the delivery of public services to the development. Impact fees have become very widespread in the United States. While the level of the fees may be prespecified, in most cases the actual fees paid by a developer are the outcome of negotiation between the city and the developer.

You have probably heard the sorry story of California's Proposition 13. A property tax revolt led to Proposition 13, which imposed a ceiling on the property tax payable on a property of 1 per cent of the most recent sales prices (with an upward adjustment of 2 per cent per year, as well as increases for improvements). This generated a fiscal crisis for local governments, which collectively agreed to give the state their property tax revenue in exchange for equalised state funding for K–12 education. One result has been a steady deterioration in the quality of K–12 education (the influx of Mexican-Americans is another cause); another has been that local governments are left with little room for budgetary manoeuvre. Fees and fines, including impact fees, have increased sharply (in Riverside, the fine for not fully stopping at a red light on a right turn is $500). I bought an older home for which the effective property tax rate at the time of purchase was slightly above 1 per cent. If I had bought a new home, the effective tax rate, including impact fee payments, could have been as high as 3 per cent. Thus, impact fees are substantial in California.

Should impact fees be encouraged or discouraged? Unfortunately, I know of only one paper that models impact fees reasonably well, and its treatment is far from definitive (Brueckner 1997). There are three general ways of financing the

infrastructure for a new development: from current revenues, through bond financing and from impact fees. Ascertaining which is best is difficult, since several separate issues are involved:

1. In the first best (that is, when there are no other distortions present), which of the methods results in the efficient timing and density of development, and in efficient migration?
2. Suppose, for the sake of argument, that individuals are infinitely lived, that a city's population keeps growing by natural reproduction, and that all the houses in new subdivisions are bought by first-time homebuyers. Does it make a difference how the infrastructure for the new subdivisions is financed?
3. Suppose instead that there are overlapping generations. Not only current residents but also future residents will occupy a new housing unit. Does it make a difference how the cost of financing infrastructure for the new subdivisions is split across generations, and, if so, why and how?
4. How do existing distortions, such as those due to property taxation, affect the comparison of the three financing modes?

6.11 Urban transportation

There is a broad consensus among urban transportation economists that mass transit is characterised by significant economies of scale. The economies of scale derive not from the technology of operations and infrastructure construction, but from *economies of service density and service frequency* (Mohring 1972). Consider the effects of doubling the number of bus travellers and doubling the density of routes, holding service frequency and fares fixed, and suppose for the sake of argument that total fleet and operating costs double. Since travellers have to walk less far on average to get to a bus stop, the average full price (time and money costs) of a trip falls. An analogous argument applies if instead service frequency is doubled and the density of routes is held fixed. Since travellers have to wait a shorter time for a bus, the average full price again falls. A recent article in the *American Economic Review* (Parry and Small 2009) documents the quantitative importance of these economies of scale. They do not alter the appropriate method of analysis for investment in mass transit.

These economies of scale may result in multiple local optima. In the Los Angeles metropolitan area, only about 2 per cent of trips and only 1 per cent of trip-miles are taken by mass transit. Service density and service frequency are low, and, because of a half-century of freeway building and federal subsidisation of auto travel, population densities are low compared to metropolitan areas of comparable size. In

such a setting it is very difficult to substantially increase the mass transit modal share. The State of California is certainly trying, however. About one third of the state's transportation budget over the past 20 years has gone to mass transit. But even with this level of investment, and despite the heavy freeway congestion (because the recession has hit California particularly hard, Los Angeles is no longer the city with the greatest number of average hours of congestion delay per auto commuter; Texas Transportation Institute 2010), it is very unlikely that the travel behaviour of Los Angelinos will be much changed.

The mass transit share is significantly higher in Australian cities. Nevertheless, it will still be difficult and costly to substantially increase the mass transit modal share. The mass transit corridor development approach, which allows an individual to lead a convenient routine without driving, seems the way to go, since it exploits the economies of scale inherent in mass transit.

Our understanding of the macroscopics of traffic congestion has increased substantially in recent years, thanks primarily to ongoing work being done at the Institute for Transportation Studies at the University of California, Berkeley. Broadly speaking, freeway traffic behaves much as the simple bottleneck model (Vickrey 1969) assumes (Cassidy and Bertini 1999). Despite the stop-and-go nature of traffic flow, at high traffic densities the average flow along a section of freeway is close to capacity. In contrast, downtown traffic congestion can be 'hypercongested', in the sense that, at high densities, average flow falls as density increases (Geroliminis and Daganzo 2008). This points to the critical importance of implementing downtown traffic management policies that eliminate recurrent high traffic density and deal effectively with high density due to non-recurrent events (such as traffic accidents, inclement weather and road construction). Cordon/area tolling has been implemented successfully in London and Stockholm, and its implementation is being actively considered in several other cities.

6.12 Build thy house upon a rock

A regular feature on the nightly news is some disaster for which a state of emergency has been declared. Disaster victims whose homes have been damaged or destroyed almost invariably receive assistance (or at least promises of assistance) from the various levels of government. But in most cases, the victims' homes would not have been destroyed if they had not built on a flood plain, or had built their homes to withstand an earthquake, hurricane or tornado. And if they had fully insured, they would have been able to get adequate compensation to build a new home. When I moved to California, I was advised *not* to purchase earthquake insurance on my home since the reduction in expected compensation I would get

from the government in the event of an earthquake would more than offset what I would get from a private insurance company after the deductible.

Private insurance against disasters is often expensive because, unlike many accidents, the risk associated with disasters tends to be systematic. Private insurers can sell some of this systematic risk on reinsurance markets, but reinsurance markets are highly imperfect. In some states, private insurers refuse to provide certain types of insurance at any price. Here is a clear instance of market failure. The federal government should intervene by providing disaster insurance itself at actuarially fair rates.

6.13 The adverse effects of macroeconomic instability in real estate markets

I live in Riverside, which is about 50 miles southeast of Los Angeles. The city and the county are about 50 per cent Hispanic. Property prices have fallen to about half their level in 2007, and below replacement cost. Since construction is the main industry in the county, the effects on the local economy have been devastating. Unemployment rose from a low of about 4.5 per cent in early 2006 to a high of over 15 per cent. In Riverside, on average about one property per block has been foreclosed.

The major cause was excessively generous mortgage underwriting. Republicans allowed it to happen because they are opposed to financial regulation, and Democrats because they want the poor to live in owner-occupied housing. Another cause was the underpricing of teaser and sub-prime mortgages in the secondary mortgage market. The episode has taught us painful yet valuable lessons. One is that mortgage markets need to be prudentially regulated (Canadian housing markets did not experience the same problem). Another is that homeownership is not necessarily the best option for poor households, since it exposes them to so much risk.

6.14 Metropolitan computable general equilibrium models

Getting an urban policy adopted entails persuading a majority of the interested parties, including voters directly or indirectly via politicians, that the policy is in their interest. This entails quantifying the effects of the policy on different segments of the population. Computable general equilibrium (CGE) models are particularly useful for this purpose because they potentially take into account all of the many

channels through which a policy affects a metropolitan area, and permit as much detail as the available data and computational speed permit.

CGE models are widely used in tax and trade policy analysis, but their use in metropolitan planning is in its infancy. Metropolitan CGE models have the same conceptual structure as other types of CGE models, except that the land and property markets at different locations are of central importance. Also, because of the durability of structures and infrastructure, and because homeowners attach so much importance to the market value of their homes, dynamics are of central importance, too.

In the past few years I have been participating in a team (Alex Alas, University of Buffalo; Michael Goodchild, University of California, Santa Barbara; and Rick Peiser, Harvard University, are the other principals) that is developing a CGE model of the Los Angeles metropolitan area, called LA-Plan.[2] The model focuses on forecasting land use, transportation and environmental quality.

The model is constructed on the basis of deterministic urban dynamic, general equilibrium theory under perfect foresight. In each period, taking the infrastructure and the stock of structures by location as given, the model solves for a temporary competitive equilibrium in which prices adjust to clear all markets, including markets for the different property types at each location. Individuals' probabilistic (due to idiosyncratic tastes) choices of residential location, work location, shopping location and travel mode, as well as their demand functions, are based on utility maximisation. Firms' probabilistic (due to idiosyncratic costs) choice of location, as well as their net supply functions (which include their input mixes), are based on profit maximisation. Between periods, developers make probabilistic (due to idiosyncratic costs), profit-maximising conversion decisions (for example, constructing at a particular density on vacant land, upgrading a building's quality, adding on, demolishing an existing structure and redeveloping at a different density). Zoning restricts the set of allowable conversions. In each period, the asset values of vacant land and property by structure type equal the expected discounted value of net rents. Thus, the model's theory is sophisticated and fully grounded in standard microeconomic theory.

The practical implementation of the model is a mammoth undertaking. First, there is data collection, cleaning and documentation, and design and implementation of the GIS data management system. We have property tax assessment data for every parcel of land on the assessment rolls in the Los Angeles metropolitan area, with information on property characteristics, as well as public census data, and several

[2] www.la-plan.org.

other databases. Second, there is imputation of missing or unreliable data. Third, there is aggregation. To keep computation times manageable, it is necessary to aggregate space into zones (at present about 100), the real transport network into an aggregated network, industries into industry groups, and households into household groups. Fourth, account must be taken of Los Angeles' geography, which renders much land undevelopable because it is mountainous, has insufficient water or is off limits for development (Indian reservations, military bases, environmentally protected areas and corridors). Fifth, the model must be calibrated. There is considerable experience to draw on in the calibration of the transport network (traffic network equilibrium models) and of travel demand, but little to draw on in calibrating the model to conform to base-year rents and values. Sixth, there is the design and implementation of a graphical, user-friendly interface.

We plan to have a preliminary version of the model up and running by the end of the summer. Then, over the next two years, we shall be running sample policy simulations, refining the model and its calibration, and developing and testing the interface.

The model builds on a series of computable, urban general equilibrium models developed by Alex Anas over the past 30 years, most of which have been implemented for Chicago. A complete description of an earlier version of the model, RELU-TRAN, is given in Anas and Liu (2007).

There are other metropolitan simulation models, such as UrbanSim and PECAS, but they are disequilibrium models in the Lowry tradition (Lowry 1964) that lack strong microeconomic foundations. God did not design the world to conform to microeconomic theory. At the same time, models that lack strong microeconomic foundations tend to be both *incomprehensible* (or, perhaps more fairly, incompletely comprehensible), in the sense that their output is hard to explain intuitively, and *incoherent and inconsistent*, in the sense that different elements of the model have different and inconsistent conceptual foundations.

Metropolitan simulation models are useful in structuring policy discussion. Their output suggests which effects of a policy are quantitatively important and which are not, as well as which groups are helped by a policy and which groups are hurt by it. Also, comparing a model's output with different sets of parameter values suggests which factors the effects of a policy are sensitive to. Their use in the context of metropolitan policy analysis has started only recently but will become more widespread.

6.15 Conclusion

This paper started by posing the questions: What problems related to metropolitan spatial growth are likely to be encountered in Australian cities in the 21st century? Anticipating these problems, how should the federal, state and local governments deal with them?

The answers provided by the paper are generally conservative. Space does not alter economic analysis in any fundamental way. The market allocates space reasonably efficiently, when the classical market failures (externalities, public goods and increasing returns to scale) are dealt with appropriately through government intervention. As in other areas of economics, if economic agents face the social costs of their economic decisions and derive the social benefits from them, they will make socially efficient decisions. This can be achieved by 'getting the prices right', which entails internalising externalities and resisting political pressures to distort prices so as to favour particular interest groups.

Non-economists tend to underappreciate the flexibility and wisdom of the market. When I was a boy, the demographic boom was in full swing. The prophesiers of doom foresaw a world with standing room only and cities clogged by traffic congestion. But economic incentives led to the demographic transition, and the large metropolitan areas responded to population and income growth through decentralisation and subcentring. As long as the economic incentives are appropriate, which is achieved by getting the prices right, urban growth will be sustainable, in the sense that at least the material quality of life will continue to improve.

This is not to say that we should just sit back and let the market do its thing. Most of the classic externalities, such as those discussed by Coase (1960), are spatial in nature. Some land uses are indeed incompatible, and those incompatibilities are appropriately dealt with by zoning, although less rigid zoning than was applied over most of the previous century. Also, if individuals are to make socially efficient decisions concerning how much to travel, what mode to travel by, where to purchase their home, and how large a home and lot to buy, they need to face the right prices, which requires internalising traffic congestion externalities. This does not require full-blown electronic congestion pricing. Getting the prices almost right will be good enough, and that can be achieved through the right combination of a gas tax and subsidies to mass transit, and perhaps cordon tolling to deal with downtown traffic congestion.

The broadly conservative theme of this paper is subject to two qualifications. The first is that spatial agglomeration derives primarily from external economies of

scale, and it is well known that the standard results concerning the efficiency of markets do not carry through in the presence of increasing returns to scale. The economic forces that give rise to cities, as well as the location of production within cities, lead to a spatial pattern of economic activity that is not fully efficient. However, the distortions are so subtle and complex that, given the current state of knowledge, we cannot say in what ways the equilibrium spatial pattern of economic activity differs from the optimum. My recommended response to this ignorance is to leave well enough alone. At the same time, we should acknowledge that planners are not necessarily wrong-headed in trying to modify the spatial pattern of metropolitan growth through land use and transportation planning.

The second qualification is that metropolitan growth may be sustainable but lead to allocations that are objectionable in some ways. A disturbing trend, at least in US cities, has been increasing residential spatial segregation by income. To some extent this trend is the outcome of natural economic forces, but it has been accentuated by exclusionary zoning in suburban communities. Even if it were completely the result of market forces, spatial segregation by income would still be objectionable, since it would increase social stratification and undermine equality of opportunity. I would be more than willing to sacrifice some efficiency to achieve greater income mixing.

References

Alonso, W. 1964, *Location and Land Use: Towards a General Theory of Land Rent*, Harvard University Press, Cambridge, MA.

Anas, A. and Liu, Y. 2007, 'A regional economy, land use, and transportation model: formulation, algorithm, design, and testing', *Journal of Regional Science*, vol. 47, pp. 415–55.

Anas, A., Arnott, A. and Small, K. 1998, 'Urban spatial structure', *Journal of Economic Literature*, vol. 36, pp. 1426–64.

Angel, S., Parent, J., Civco, D. and Biel, A. 2010, 'The persistent decline in urban densities: global and historical evidence of sprawl', Lincoln Institute of Land Policy Working Paper.

Arnott, R., Atkinson, A.B., Arrow, K. and Drèze, J.H. (eds) 1994, *Public Economics Selected Papers by William Vickrey*, Cambridge University Press, Cambridge, United Kingdom.

Bairoch, P. 1988, *Cities and Economic Development: From the Dawn of History to the Present*, University of Chicago Press, Chicago.

Bator, F. 1958, 'The anatomy of market failure', *Quarterly Journal of Economics*, vol. 72, pp. 351–79.

Blakely, E. and Snyder, M. 1997, *Fortress America: Gated Communities in the United States*, Brookings University Press, Washington DC.

Brueckner, J.K. 1987, 'The structure of urban equilibria: a unified treatment of the Mills–Muth model, in Mills, E.S. (ed.), *Handbook of Regional and Urban Economics 2*, North-Holland, Amsterdam, pp. 821–45.

—— 1997, 'Infrastructure financing and urban development: the economics of impact fees', *Journal of Public Economics*, vol. 66, pp. 383–407.

Burchfield, M., Overman, H., Puga, D. and Turner, M. 2006, 'Causes of sprawl: a portrait from space', *Quarterly Journal of Economics*, vol. 121, pp. 587–633.

Cassidy, M. and Bertini, R. 1999, 'Some traffic features at freeway bottlenecks', *Transportation Research B*, vol. 33, pp. 25–42.

Clark, C. 1951, 'Urban population densities', *Journal of the Royal Geographical Society*, vol. 114, pp. 490–96.

Coase, R. 1960, 'The problem of social cost', *Journal of Law and Economics*, vol. 3, pp. 144–71.

Cronon, W. 1991, *Nature's Metropolis: Chicago and the Great West*, Norton, New York.

Debreu, G. 1959, *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*, Yale University Press, New Haven, Connecticut.

Duranton, G. 2009, 'Are cities engines of growth and prosperity for developing countries?', in Spence, M., Annez, P. and Buckley R. (eds), *Urbanization and Growth*, World Bank, Washington DC.

Fujita, M. and Ogawa, H. 1982, 'Multiple equilibria and structural transition of non-monocentric urban configurations', *Regional Science and Urban Economics*, vol. 12, pp. 161–96.

Fujita, M., Krugman, P. and Venables, A. 1999, *The Spatial Economy: Cities, Regions, and International Trade*, MIT Press, Cambridge, MA.

Fujita, M. and Thisse, J.-F. 2002, *Economics of Agglomeration*, Cambridge University Press, Cambridge, United Kingdom.

Geroliminis, N., and Daganzo, C. 2008, 'Existence of urban-scale macroscopic fundamental diagrams: some experimental findings', *Transportation Research B*, vol. 42, pp. 759–70.

Glaeser, E. and Kahn, M. 2001, 'Decentralized employment and the transformation of the American city', Working Paper 8117, National Bureau of Economic Research.

Glaeser, E. and Kohlhase, J. 2003, 'Cities, regions, and the decline of transport costs', *Papers in Regional Science*, vol. 83, pp. 197–228.

Glaeser, E., Gyourko, J. and Saks, R. 2005, 'Why have housing prices gone up?', *American Economic Review*, vol. 95, pp. 329–33.

Gordon, P. and Richardson, H. 1986, 'The distribution of population and employment in a polycentric city', *Environment and Planning A*, vol. 18, pp. 161–73.

——, —— 1989, 'The influence of metropolitan spatial structure on commuting time', *Journal of Urban Economics*, vol. 26, pp. 138–51.

——, —— 1996, 'Beyond polycentricity: the dispersed metropolis, 1970–1990', *Journal of the American Planning Association*, vol. 62, pp. 289–95.

Hamilton, B. 1975, 'Zoning and property taxation in a system of local governments', *Urban Studies*, vol. 12, pp. 205–11.

Harrington, M. 1962, The Other America, McMillan, New York.

Harris, J. and Todaro, M. 1970, 'Migration, unemployment and development: a two-sector analysis', *American Economic Review*, vol. 60, pp. 126–42.

Henderson, J.V. 1974, 'The sizes and types of cities', *American Economic Review*, vol. 64, pp. 640–56.

—— and Thisse, J.-F. (eds) 2004, *Handbook of Regional and Urban Economics 4: Cities and Geography*, Elsevier, Amsterdam.

Kanemoto, Y. 1980, *Theories of Urban Externalities*, North-Holland, Amsterdam.

Kaplow, L. 2008, *The Theory of Taxation and Public Economics*, Princeton University Press, Princeton, New Jersey.

Kozol, J. 1967, *Death at an Early Age*, Houghton Mifflin, Boston, MA.

Lowry, I. 1964, *A Model of Metropolis*, Research memorandum RM-4035-RC, Rand Corporation, Santa Monica, CA.

Madden, J. 2003, 'The changing spatial concentration of income and poverty among suburbs of large US metropolitan areas', *Urban Studies*, vol. 40, pp. 481–503.

Mieszkowksi, P. and Mills, E. 1993, 'The causes of metropolitan suburbanization', *Journal of Economic Perspectives*, vol. 7, pp. 135–47.

Mills, E. 1967, 'An aggregative model of resource allocation in a metropolitan area', *American Economic Review*, vol. 57, pp. 197–210.

Mirrlees, J. 1972, 'The optimum town', *Swedish Journal of Economics*, vol. 74, pp. 114–35.

Mohring, H. 1972, 'Optimization and scale economies in urban bus transportation', *American Economic Review*, vol. 62, pp. 591–604.

Muth, R. 1970, *Cities and Housing*, University of Chicago Press, Chicago.

Papageorgiou, Y. and Pines, D. 1999, *An Essay on Urban Economic Theory*, Kluwer, Boston, MA.

Parry, I. and Small, K. 2009, 'Should urban transit subsidies be reduced?', *American Economic Review*, vol. 99, pp. 700–24.

Saiz, A. 2010, 'The geographic determinants of housing supply', *Quarterly Journal of Economics*, vol. 125, pp. 1253–96.

Salanié, B. 2000, *Microeconomics of Market Failures*, MIT Press, Cambridge, MA.

—— 2003, *The Economics of Taxation*, MIT Press, Cambridge, MA.

Solow, R. 1972, 'Congestion, density and the use of land in transportation', *Swedish Journal of Economics*, vol. 74, pp. 161–73.

Smith, A. (1776) 2009, *The Wealth of Nations*, Classic House Book, New York.

Strange, W. 1995, 'Information, holdouts, and land assembly', *Journal of Urban Economics*, vol. 38, pp. 317–32.

Texas Transportation Institute 2010, *2010 Urban Mobility Report*.

Tiebout, C. 1956, 'A pure theory of local public goods', *Journal of Political Economy*, vol. 64, pp. 416–22.

von Thünen, J. (1827) 1966, *The Isolated State*, trans. C.W. Wartenberg, Pergamon Press, Oxford.

Vickrey, W. 1969, 'Congestion theory and transport investment', *American Economic Review*, vol. 59, pp. 251–60.

Warner, S. 1962, *Streetcar Suburbs*, Harvard University Press, Cambridge, MA.