28/2/18

Place-Based Policies for Development

Gilles Duranton

Anthony J. Venables

Wharton School

University of Oxford

Abstract

Many development policies, such as placement of infrastructure or local economic development schemes, are "place-based." Such policies are generally intended to stimulate private sector investment and economic growth in the treated place, and as such they are difficult to appraise and evaluate. This paper sets out a framework for analyzing the effects of such policies and assessing their social value. It then reviews the literature on place-based policies in the contexts of transport improvements, economic corridors, special economic zones, lagging regions, and urban policies.

JEL classification: O10, O18, R10, R11, R13
Keywords: Placed based policies, spatial regions, economic corridors, lagging regions, urban.

Acknowledgements: This paper was prepared for the World Bank under a program on spatial productivity, supported financially by UK DFID. The authors have had a series of meetings with World Bank staff working on place-based policies, and thank them for their time and practical insights. Thanks also to Bill Maloney, Somik Lall and participants in the Space and Productivity conference held in the World Bank, September 2017, for valuable comments.

Authors' addresses:
G. Duranton,
The Wharton School,
University of Pennsylvania,
3620 Locust Walk
Philadelphia, PA 19104.
USA
duranton@wharton.upenn.edu

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

Place-Based Policies for Development

Table of Contents:

1. Place-based policies: Introduction

Place-based policies – such as transport improvements, special economic zones, and treatment of lagging regions – are an integral part of development policy. They are frequently very expensive and they require that choices be made between the places at which policy is implemented. Choices are particularly stark in developing countries as constrained resources and the lumpiness of many public investments mean that, of necessity, some places will be served with roads, telecommunications, power and other public services, before others. Informing these difficult choices with systematic ex ante appraisal of projected policies is complex because of two distinctive features of place-based policies.

The first is that place-based policies are typically expected to deliver two types of benefit. One is the direct effect; e.g. a transport improvement saves time and vehicle operating costs. The other is indirect; the transport improvement may induce private sector investment, job creation, and higher productivity.[1] The direct effect is the stuff of standard project appraisal techniques for which well-established techniques are in place, based around cost-benefit analysis. But indirect effects depend on responses of the private sector, often including major location decisions that incur sunk costs and are long-lasting. Such decisions and their economic impacts are hard to predict.

The second reason is inherent in the spatial context. The value of proximity – and the cost of distance – creates a high degree of complementarity in the location of investments, public and private. On the public side, having good roads but no electric power leaves a place unattractive for private investors; concentrated packages of investment matter. Private investment is encouraged by public infrastructure and also by investments made by other private investors – the location decisions of workers, customers and supplier firms. Expectations about the future development of a place are critical in shaping these decisions. These factors can create spatially focused increasing returns to scale (agglomeration economies) and virtuous or vicious cycles of development, yielding spatially uneven outcomes.[2] They create numerous market failures so there is a role for policy, but it is these factors that make the effects of policy particularly difficult to predict.

These two features pose a major challenge for appraisal of place-based policies (PBP). A conventional cost-benefit analysis would simply ignore indirect effects, but this is incorrect if – as is likely in the spatial context – there are important market failures. Ignoring indirect effects may lead to the unsatisfactory situation in which the strategic case for a project is presented in terms of these effects, while they are ignored in the narrow economic appraisal. Including them

[1] Some literature refers to 'wider benefits'; we will use the term 'indirect effects' throughout.

[2] See World Bank (2009).

risks opening the floodgates to poorly grounded (and probably over-optimistic) claims about both their magnitude and their social value.

This paper sets out a framework for addressing the impact of PBP in a systematic and rigorous manner. The framework is broad enough to be relevant and adaptable to a variety of circumstances, yet rigorous enough to test, evaluate, and frame quantitative estimates of the arguments put forward. The bulk of the paper is devoted to the economic issues – in general and in specific contexts – to do this. We start with a broad summary of the seven principles that we think need to enter an economic appraisal of a proposed PBP.[3]

1. Narrative: There should be a clear narrative of the main problem that policy is intended to address and the key market failure(s) that motivate the policy.

2. Quantity change: There should be a full description of the expected ***quantity changes*** arising from the project, including further economic activity likely to be created by the policy. This should separate clearly the direct and indirect effects, i.e. distinguish between the goods and services delivered and used by the project, and quantity changes due to induced changes in private sector behavior. It should be accompanied by a description of the mechanisms through which these quantity changes may arise. The analysis must take into account possible displacement effects, i.e. policy induced relocation of economic activity between places.

3. Valuation of changes: There should be analysis of why quantity changes brought about by the project are of net social value, distinguishing between the value of direct and indirect effects. The latter may be of net social value because of the interaction of quantity changes with market failures and inefficient resource allocation; the magnitude of and reasons for any such market failures should be carefully diagnosed and evaluated. Net social value may also derive from equity concerns, which should be made explicit.

4. Transparency: The mechanisms underpinning both the quantity changes and their social value should be clear, and should be explained in a manner that enables the key magnitudes to be understood from straightforward back-of-the-envelope calculation.

5. Sensitivity: There should be analysis of the dependence of the quantity effects and their valuation on key assumptions about the economic environment. Scenarios outlining the quantitative importance of failure of these assumptions should be outlined.

6. Complementary policies: There should be a thorough consideration of complementary measures that are needed for successful implementation of a project. While we advocate that projects should have a simple and clear rationale, we understand that the success of PBP may depend on many factors and in many cases, we may be deep into the *n*-th best.

[3] The bulk of this paper focuses on the second and third principles in our list because they are analytically the most challenging and the other principles derive directly from them.

7. Alternatives: Any project should make a strong case that it provides the most cost-effective way to solve the main problem described in the narrative. For instance, accessibility can be improved by providing faster mobility, but it can also be improved by providing greater proximity. If the latter is cheaper, why engage in the former?

2. Place-based policy: Principles

Systematic analysis of the effects of policy has two components. The first is to establish the ***quantity effects*** of the policy; what changes in levels of employment, output, or other variables of interest are brought about by the policy compared to a 'do-nothing' alternative? The second is ***valuation*** of the quantity changes; are there reasons to believe that an increase in outputs – and the associated increase in use of inputs – is of social value?[4] Each of these requires an understanding of the response of the economy to change, and of the inefficiencies and market failures that are present in the economy. The spatial context is distinctive in both these respects, and we start with a brief overview of key elements of spatial economics (section 2.1).

The framework we develop for analyzing PBP is based on breaking policy design down into its two constituent parts of quantity change and valuation, and is developed in section 2.2. Its application is highly context specific, and remaining sections of the paper discuss a number of areas of application. Section 3 applies the framework to long-distance transport and economic corridors, special economic zones, and lagging regions. Section 4 turns to the urban context and discusses urban transport and housing.

2.1. The spatial context

Standard economic analysis is based on diminishing returns – the more of an activity is undertaken, the lower is the value of doing still more. Applying this in a spatial context yields the prediction that economic activity will be smeared more or less smoothly across space, a prediction contradicted by the existence of towns and cities. To think spatially, several ingredients have to be added.

First, ***proximity*** is valuable, which implies that economic activity has a tendency to cluster and is unevenly distributed across space. There are two drivers of this. One is the direct efficiency saving of being close together. There are savings in transport and communication costs, and economies of scale arise as firms and infrastructure operate at scale. There is evidence that

[4] We do not address the issues surrounding construction costs, finance and procurement of projects.

provision of infrastructure is much cheaper, per person, in dense urban areas than dispersed rural ones.[5]

The other driver is a set of ***agglomeration economies*** that are generated by close and intense economic interaction. These arise through several different mechanisms. Thick labor markets enable better matching of workers to firms' skill requirements. Better communication between firms and their customers and suppliers enables knowledge spillovers, better product design and timely production. A larger local market enables development of a larger network or more specialized suppliers. Fundamentally, larger and denser markets allow for both scale and specialization.

A good example is given by specialist workers or suppliers. The larger the market the more likely it is to be worthwhile for an individual to specialize and hone skills in producing a particular good or service. The presence of highly specialized skills will raise overall productivity. The specialist will be paid for the product or service supplied but, depending on market conditions, is unlikely to capture the full benefit created.[6] Since the benefit is split between the supplier and her customers there is a positive externality. And this creates a positive feedback – more firms will be attracted to the place to receive the benefit, growing the market, further increasing the returns to specialization, and so on. This is the classic process of cluster formation.

The benefits of proximity are a force for increasing returns, but proximity also has ***costs***, particularly in the urban context, creating a force for decreasing returns. Clustering of firms increases commuting costs for workers who may have to travel far to employment centers such as the central business district. These costs are exacerbated by congestion – a negative externality – and other costs of close urban living. Land becomes the scarce factor and housing consumption (floor-space per household) is reduced. However, it is important to note that high land rents and land prices are not a real cost; they are a transfer payment from occupants of land to its owners, so do not use up real resources (as does time in commuting).

Location choices – be they by firms or by households – are typically major decisions, with large sunk costs and, if structures are being built, creating long-lived assets. ***Expectations*** of future returns are therefore critical, and agglomeration economies mean that the returns to investing in a place depend on who else is (or is expected to be) there. This in turn creates a ***first-mover***

[5] For example, estimates contained in Foster and Briceno-Garmendia (2010) suggest that the per capita capital cost of providing a package of utilities and infrastructure may be five times higher in a low-density city than a high-density city, and ten times higher in rural areas (Foster and Briceno-Garmendia 2010, Table 5.6).

[6] The supplier will capture the full benefit only if able to perfectly price discriminate. Otherwise, the customer will also receive some consumer / user surplus on the introduction of a new product. This observation is central to the wide range of economic models in which the number and variety of goods and services offered is endogenous (following Dixit and Stiglitz 1977).

problem: no one wants to move to a new place while uncertain about its future development. There is inertia and path dependency as firms are unwilling to move out of existing clusters, and it is therefore hard to start new ones. Coordinated movement by a group of firms might solve the problem, but generally, there is ***coordination failure***. Thus, a planner can construct a model of a perfectly functioning new town but there has to be a path of public and private sector investment and location decisions that lead from the initial situation of empty fields to the completed town. If such a path is not in place, then development will fail. These arguments are particularly acute in a developing country. This is partly because the economic environment is one of rapid change in all respects, including spatial. And also because the cost of making a poor location decision may be higher. In a developed economy, even if a firm does not have a supplier of some specialist component in the same city, there is probably a supplier within 24 hours' delivery time. In a developing country this is not true, so the need for coordination – and the cost of coordination failure – become much greater, at least in relative terms.

A further element of spatial economics concerns the role of the ***price and wage adjustment*** in securing regional competitiveness. Coordination failure makes it difficult to establish alternative economic centers, but why doesn't the price mechanism make the alternative place so cheap that firms will move there? The answer is that, within a country, the performance of a region depends largely on its absolute advantage, not its comparative advantage. If a *country's* export sector has a negative shock the adjustment mechanism is a real depreciation, i.e. a reduction in its wage and unit costs relative to its trading partners, and this reduction continues until other sectors become competitive. If a *region* within a country suffers a negative shock there may be little flexibility of relative wages between regions as labor markets are relatively more integrated. Integration is part institutional, and part due to intra-country mobility of labor and capital that will tend to equalize real wages across regions. Of course, there are some immobile factors, such as land and houses. Their prices will fall in the adversely affected region but since these factors only represent a small fraction of business costs they have little leverage in bringing sectors to the point of competitiveness.[7] Since the mechanism of real depreciation is largely absent, it follows that regional inequalities are likely to be persistent, and that the market response is not to move new employment into the area, but to (eventually) move labor out.

[7] That immobile factors cannot adjust to induce a real depreciation that fully offsets a negative shock does not mean that immobile factors play no role. The durability of immobile factors implies that they slow down the movement of mobile factors as argued by Glaeser and Gyourko (2005). After a negative local shock, the demand for immobile factors, housing and commercial properties, declines. Because their supply is inelastic in case of a negative shock (housing cannot be profitably un-built), the price of immobile factors can fall essentially to zero. This induces some workers to stay because of a big drop in the cost of living locally. Then, because housing and commercial properties are so cheap, they are not worth investing into and are left to decay. It is only as the stock of properties slowly decays that mobile factors fully adjust. This process is of course particularly slow and generates additional issues of its own as the stayers tend to be those with the weakest opportunities elsewhere.

These factors combine to produce a distribution of economic activity across space, and this pattern is typically highly uneven. There is a distribution of cities and urban areas with different specialisms and of different sizes, and regions with different levels of activity and income. Three important points follow.

First, the market outcome (such as the size of cities or the number and location of clusters) is not, generally, efficient. In the urban context, private decisions to locate in a city mean that, for the marginal entrant, private benefits are equal to private costs. But external effects created are not taken into account in this decision so there may be net social benefits (or costs) from expanding a city according to whether or not the value of positive external benefits created is greater (or less) than the cost of negative ones. In particular, coordination failure means that it is hard to start new centers of activity and therefore there may too few such centers. Thus, an important consequence of coordination failure is excess primacy, meaning the tendency for the largest cities in a country to be excessively large compared to those further down the urban hierarchy. We know from the data that this is a common feature of developing countries, compared to higher income countries now and also when they were at comparable stages of development (Kim, 2008).

Second, while the spatial pattern of activity is not 'first-best' efficient, outcomes with agglomeration – such as a city – yield an aggregate net benefit. The reason is essentially that part of the costs faced by (intra-marginal) members of the cluster is land rent which, as we saw above, is a transfer payment not a real cost. Thus, in the simplest setting, all the land value appreciation that occurs in a city or in a cluster is the capitalized flow of the net benefit created by the city. If land values in larger cities are higher than those elsewhere, then we can be confident that their productivity advantage means that they are producing aggregate net benefit.[8]

Third, the aggregate benefits of agglomeration combined with inefficient market outcomes provide a strong justification for PBP, and it is useful to distinguish between lagging and leading PBPs. Much PBP lags development, responding to bottlenecks such as congestion. Economic activity has then revealed where investment is needed, and this reduces uncertainty about the effects of policy. But lagging policy has disadvantages, putting the economy through a period when constraints are costly and incurring the costs of retrofitting infrastructure in crowded places. Alternatively, PBP may lead development in a particular place, potentially acting as a coordination mechanism. As examples, an infrastructure project may be a credible commitment that a place is selected for development, and construction of transport lines creates a focal point for development in a growing city. This can shape expectations and act as a catalyst to trigger private sector investment. Some PBPs are intended to achieve these effects, but such policies are inevitably more speculative than those responding to existing levels of activity.

[8] These arguments are developed more fully in section 4.

2.2. Policy: Quantity change and valuation

Economic policy has both direct and indirect quantity effects. Direct effects arise as e.g. building a hospital will improve health outcomes. Indirect or induced effects occur as building a road will not just increase the number of journeys made, but may also increase private investment as places along the road become more attractive locations for economic activity. PBP is unusual in that its principal objective is often the indirect effect, and these are inevitably difficult to predict with confidence. This is particularly so as the desired indirect effects typically encounter the issues discussed in the preceding sub-section.

The following framework is useful. Think of the economy as being represented by some general production function, $Y = AF(x, L, K)$, where Y is output (or, ultimately, social welfare). This depends on the policy, x, and also on labor and capital in the economy. L and K denote these factors across all uses and places (i.e. they are vectors). The function F is the economy's production function and A denotes a productivity factor, again, to be interpreted as different across sectors and places. The direct effect of the policy is the change in x and its impact on Y, given L, K and A. Spatial policy is typically looking also for indirect effects, induced changes in the location and levels of activity, L and K, and in productivity A. Together these are the ***quantity effects***, and the first stage of policy design and appraisal is to establish these quantity effects.

The second stage is to place a value on these quantity changes. They will generally create costs and benefits, and ***valuation*** has to work out the net effect. The value of a change in total factor productivity may be clear, but changes in L and K may be principally a spatial shuffling, creating investment in one place at the expense of another. Valuation has to compare the values in each place, to establish whether the net effect is positive or negative. In short, valuation requires establishing both the value of a factor in its new use, and the opportunity cost of this employment.

i) Quantity effects: Induced private investment

To establish the form and size of quantity effects requires understanding two things. One is what drives private sector location and investment decisions, principally those of firms but also the decisions of workers. The other is displacement, i.e. the general equilibrium response of the economy; this determines what other sectors or areas might contract in response to growth in a particular place. Both of these are context specific, and different contexts and polices are discussed in sections 3 and 4 of this paper. However, a number of general points apply.

Location and investment decisions are inherently non-marginal (they are either/or choices, rather than fractional adjustments). They are costly, typically incurring sunk costs and shaped by expectations. The private sector will decide to invest in a place only if multiple conditions are

met, conditions that can be grouped into different types, such as the natural characteristics of the place, the policy environment, and the 'business ecosystem'. We focus on the last two of these.

The policy environment covers national variables and those that are place-based, i.e.

- Infrastructure including utilities, transport.
- Place-specific tax and regulation, as in a special economic zone.
- Policy as it affects labor supply, e.g. public services, commuting, housing.
- Institutions including the clarity and enforcement of property rights and contracts.

We borrow the term 'business ecosystem' from the business literature to mean the network of organizations – including suppliers, distributors, competitors, customers and workers – that contribute to the performance of firms and the value of investment decisions that they undertake in a particular place. This includes:

- Related firms: Does the place have a stock of firms and other productive activities, in particular its suppliers and customers?
- Workers: the supply of workers with appropriate skills at competitive wages, or the potential to attract migrants to the area.
- The availability of other factors, land and capital.
- Market size; the size of markets to which the place is well connected.

There are several critical points about these conditions.

First, each of the conditions is necessary for private investment to be undertaken (or if not strictly necessary, there is a high degree of complementarity between elements of the list). This creates a weakest link problem and hence threshold effects and discontinuous responses of private investment to policy levers. For example, adding more utilities in a place may have no effect if other conditions are not present; or, if other conditions are met, more utilities may push the place across a threshold and trigger a large private investment response. Quite obviously, good roads for exports will not be enough if a place lacks reliable electricity or if they do not reduce costs to a level at which firms can compete with producers from other locations. These strong complementarities make it inherently difficult to predict quantity changes.

Second, the policy environment list covers different areas of government and, to be effective, policy requires coordination across space, function, and time. Policy, and the ability to implement it effectively, needs to be integrated spatially. For example, in order to grow a city, authority is needed within the city and also in the environs into which it will grow. Policy needs to be integrated functionally, i.e. covering planning, land and building regulations, infrastructure and utility and public service provision. And policy needs a long-view, able to make credible commitments to future city development. Some coordination and consistency are needed

between the different levels of government, local, regional, and national. Having competence, financial resources, and credibility to meet these challenges requires an authorizing environment more integrated than that which is typically present between different parts of government.

Third, the business ecosystem part of the list largely mirrors the earlier discussion of agglomeration and clustering. Conditions that determine one investor's decision depend on decisions that have been – or will be – taken by other private sector agents. This extends across a wide range of agents and depends on expectations. Thus, the location decisions of firms depend on those of workers and other firms; the location decisions of workers depend on firms and on house-builders; house-builders are themselves taking a long view about employment and population in the place.

Finally, expectations are once again crucial. What matters is whether these conditions are expected to be in place over the duration of a long-lived investment.

All these points indicate different ways in which policy can shape private investment and location decisions. Some are those in the 'policy environment' list above, but others (which we will term 'soft') involve shaping expectations about the business ecosystem. For example,

- Providing credible signals that a particular place will develop, such as a city plan or commitment to future hard policies.
- Setting up an effective investment agency to work with the private sector.
- Finding a 'large' private agent that can internalize or coordinate many of the private sector actions required for a cluster to operate.
- Commitment to remove future blockages and obstacles as they arise.

Displacement and general equilibrium effects are the quantity changes that occur – possibly in quite different places – in response to changes brought about by the policy. It is not always either necessary or possible to identify these with precision but, if capital and labor are simply being reshuffled between uses, then both sides of these quantity changes have to be accounted for.

Displacement effects can occur through several distinct routes. The most direct is competition for a particular project – such as a single factory that will operate in just one of many possible places. More generally, displacement may arise through the product market in which, if demand is inelastic, an increase in supply in one place will be met by a reduction in supply elsewhere. This effect is most pronounced for non-tradable goods, where demand comes just from a local or national market. Displacement may also occur through factor markets; if there is a fixed supply of capital or full employment of a given labor force, then expansion of one activity must be accommodated by contraction of another. As we discuss below, displacement is a major issue with PBPs that can take many different forms, depending on the exact policy at hand.

ii) Valuation: Market failure

Given some predictions about the quantity changes arising from a policy intervention, what is their combined social value? Valuation of the direct effects (the effect of a change in x, given A, L, K) is the stuff of standard cost-benefit analysis. Thus, in the transport context the direct effects include the 'user benefits' of time savings, and their valuation is typically based on studies of the value of time. More generally, outputs may be valued at market prices (as they would be in a commercial decision) or, if outputs are non-marketed, values have to be inferred (e.g. by studies of the value of time, Small and Verhoef, 2007).

What about indirect effects, the shuffling of L and K between places and sectors of the economy due to induced private sector investment decisions?[9] The benchmark case is that the value of these changes is zero. If the economy is efficient then the marginal value of L is the same in all its uses, and so too is the marginal value of K. Moving them between uses is therefore of zero value – additional output in the new use is worth just the same as lost output in the alternative. This argument is the rationale for only looking at the direct effects of policy, as in standard cost-benefit analysis; if the economy is efficient, then other indirect effects combine to have zero value.

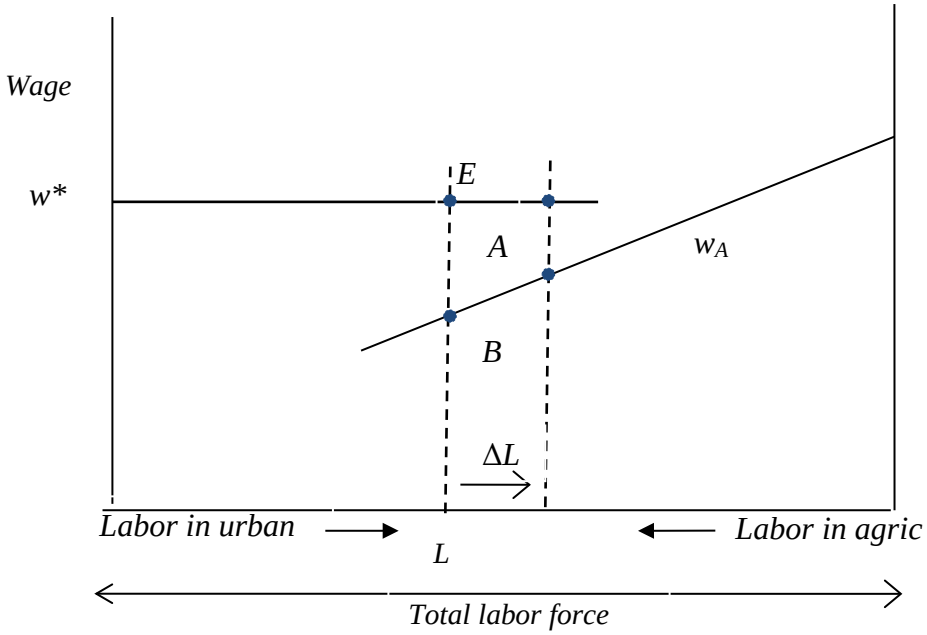
Indirect effects are of net value only if policy tends to correct inefficiencies, i.e. operates to raise productivity or to draw resources from lower to higher value use. This in turn arises only if the price-system (or some other mechanism) has not lined up marginal values. In the spatial context in particular there are many obstacles to efficient allocation, as we saw in section 2.1, and valuation requires that such inefficiencies are diagnosed, quantified, and understood. The remainder of this section outlines some of the key markets in which such failures are prevalent. They are linked to particular spatial contexts and policies in greater detail in sections 3 and 4 of the paper.

Job creation: spatial and labor market inefficiencies: Frequently the main motive for PBP is the creation of new jobs in the affected place, created as a consequence of increased investment and economic activity. Valuation of these jobs depends on the value of workers in the new activity compared to what they would otherwise have been doing. Box 2.1 illustrates the simple analytics.

[9] And perhaps also from abroad or from additional savings, but both of these at a (opportunity) cost.

----------**Box: 2.1: Labor reallocation and productivity gaps**----------

The productivity of labor in non-agriculture is between two and three times as high as it is in agriculture (Gollin et al. 2014). The implications of this for valuing new modern sector jobs are illustrated below. The length of the horizontal axis is the total labor force, divided between urban employment at productivity w^* and agricultural employment with value productivity w_A, measured from left and right axes respectively. The initial division of the labor force between sectors is at point L. What is the value of creating ΔL new urban jobs? If workers come from existing urban jobs of similar productivity then there is 100% displacement, total urban employment remains at L and no surplus is created. If they come from agriculture then agricultural output lost is B and urban output gained is A + B, giving net surplus of A.



This analysis begs the questions: Why is there a productivity gap in the first place? What determines the quantity response? For example, the classic Harris-Todaro model reconciles migration and a productivity gap by having a pool of urban un(der)-employed workers. In this model creating ΔL more urban jobs pulls more than ΔL workers out of agriculture and into urban unemployment (attracted by the probability of securing an urban job), in which case the loss of agricultural output is greater than B and the net value can be positive or negative.

To be concrete, consider an example of a regional training scheme for managers in manufacturing firms, a typical local economic development policy aimed at capacity building. This training scheme may increase the productivity of local manufacturing firms. There may be a resulting quantity effect as manufacturing employment output expands, perhaps by a lot. This does not necessarily mean that we should deem this scheme to be especially socially valuable. The value of these changes in quantity will be very low if new manufacturing workers are pulled

out of other sectors where they were nearly as productive. But if the new manufacturing workers were previously unemployed, perhaps because of some frictions on the labor market, the value of the extra jobs is high. Alternatively, the value of this change in quantity will be high if manufacturing activity generates positive externalities for economic development.

Externalities: clustering and congestion: As suggested in section 2.1, externalities are created by the location – and concentration – of economic activity. Some of these are positive (cluster/agglomeration effects on productivity) and some negative (urban crowding and congestion). As we show below, the balance between these forces can lead to clusters that are too small, because agglomeration externalities are not appropriately internalized. The balance can also lead to clusters that are too big because of failure to internalize congestion costs, and because of the difficulty of starting new clusters. This difficulty – the coordination failure and first mover problem – means that migrants continue to pile into existing clusters due to lack of alternatives. When crowding and congestion forces dominate, we expect to create value by enabling a new cluster to form. The development of a new center can take pressure off an existing one and create benefits by reducing congestion. Instead, when agglomeration forces dominate or when the income from land leaks outside, expanding extant clusters makes sense.[10]

PBP may affect city size directly (e.g. controlling flows of migrants) but is more likely to do so indirectly, by increasing urban benefits (e.g. service provision or local employment opportunities) or by reducing urban costs (e.g. transport improvement). The latter can have positive value both by directly reducing congestion and by enabling cluster expansion and hence external economies of scale and productivity growth. However, valuation of such changes has to be seen also in the context of inefficiencies in the overall city size distribution created by factors such as coordination failure and manifest in excess primacy.

Capital, land and building: Patterns of land-use and investments in buildings – commercial and residential – encounter numerous market failures. In urban areas land is the ultimate scarce factor, so there is high return to using it efficiently. However, unclear property rights and obstacles to trading land can result in inefficient land use. Building is impeded by failures in capital markets (particularly for residential mortgages) and in some cases inappropriate building and land-use regulation.

PBPs addressed at these imperfections have direct benefit if they enable land to switch from a low- to a higher-value use. Multiple effects are likely to operate. For example, a regulatory

[10] Again, one needs to be wary of displacement effects. When cluster expansion occurs because of labour relocation from other areas, agglomeration gains in the targeted area will come at the expense of agglomeration losses elsewhere. In the specific case where the agglomeration elasticity is constant, the gains in the targeted area will be exactly offset by the losses elsewhere. To understand this result, assume for instance that output per worker grows by 1% every time cluster employment grows by 10%. With a PBP policy, assume that a cluster with 100 employees grows by 10 at the expense of a loss of one employee for 10 other clusters of the same size. Output per worker in the receiving cluster grows by 1% at the expense of a loss of 0.1% in 10 other clusters of the same size. This is a wash.

change that enables efficient use of a piece of urban land might yield social benefit in three distinct and additional ways. By increasing the productivity and raising the value of land;[11] and via indirect effects that interact with other market failures, such as creating formal sector jobs (if the marginal product is greater than the alternative in the informal sector) or growing the urban cluster with positive external economies.

Human capital development and inequalities: Many of the most important externalities develop slowly through time as human (and social) capital evolves at different speeds in different places. These changes are particularly hard to quantify but are surely important. Lagging regions may lose disproportionately more skilled labor than unskilled labor, including in professions such as health and education. The negative consequences of poorer schooling and bad health outcomes in these regions may outweigh benefits in regions that attract these skills. In any case the implications for inequality – and its political consequences – create a greater positive value on expanding activity and raising incomes in poor, as compared to richer regions.

Policy targeting and the second best: Finally, we note that a particular PBP is unlikely to be the 'first-best' way to address a market failure. The theory of policy targeting tells us that the fundamental determinants of market failure should be diagnosed and then addressed by targeted policy. Thus, if unclear land rights hamper building, then the first best policy is to clarify these rights, and so on. PBP is likely to be second (or *n*-th) best policy; policy makers should be aware that better policies, more closely targeted on the market failure, may exist.

2.3. Framework summary

Our discussion has highlighted two dimensions that help to organize the appraisal of PBP. One is the distinction between direct and indirect effects. The other, the change in quantities associated with each, and the valuation of these changes. Table 1 tabulates these as an aide-memoire, recording in the body of the table some of the key issues associated with each. The bottom rows of the table emphasize displacement: the fact that changes in quantities are not all proximate to the project, yet must be taken into account in a full valuation of net social value.

With this by way of introduction, we now turn to specific place-based policies. Section 3 deals with regional issues including special economic zones, corridors and lagging regions, and section 4 focuses on urban policies.

[11] There is a tendency to use changes in land prices to value PBP. Changes in land prices will provide an unbiased measure of welfare changes locally only under restrictive conditions and will in general fail to do so when factors are imperfectly mobile, residents are heterogeneous, or lot size is endogenous (Kanemoto, 1988). We return to this issue below.

Table 1: A Framework

<table>
<tr><td></td><td>Quantity change:
- What does policy change?</td><td>Valuation:
- Are quantity changes of net social value?</td></tr>
<tr><td>Direct effects</td><td>The project:
- Supplies goods and services
- Uses inputs</td><td>- Market prices or imputed values
- Market failure & marginal values e.g.
  - Labor markets and the shadow wage rate.</td></tr>
<tr><td rowspan="2">Indirect effects</td><td>Induced changes in private activity:
- Multiple conditions ‘necessary’ for change
- Agglomeration and the business ecosystem
- Expectations and coordination failure
- Path dependence and excess primacy</td><td>Benchmark: ‘small’ quantity changes are of zero social value as MSB = MSC.
Quantity changes may interact with inefficiencies due to e.g.
- Externalities:
  - Agglomeration/ Congestion
- Lock-in to inefficient spatial patterns
  - City size/ Lagging regions</td></tr>
<tr><td colspan="2">Displacement
Relocation vs. creation of activity | NET social value and opportunity cost</td></tr>
</table>

3. Corridors, Zones and Regions

This section reviews some of the issues and literature surrounding PBPs that are designed principally to trigger new development, focusing on corridors and long-distance transport improvements, special economic zones, and lagging regions. In section 4 we concentrate on the urban context, looking at PBPs in areas with existing high levels of economic activity.

3.1. Corridors and long-distance transport improvements

The development of medium- or long-distance transport improvements typically has both the objective of reducing costs for users of the route, and of stimulating economic activity along the length of the route (as with agricultural corridors) or at towns and cities at various points. Analysis of whether these objectives are likely to be achieved by a particular project is hard. No one would doubt that a completely isolated place will be poor, or that most rich places are well connected.[12] But it does not follow from these observations that all well-connected places are rich or that improving connectivity necessarily brings development. As outlined in section 2.1, economic development is spatially uneven and prone to clustering. Changes in economic geography brought about by transport improvements might mean that some places gain, others lose.

In this section we start by looking at analytical arguments on the 'quantity effects': how might the location of economic activity be affected by transport improvement? We then turn to the empirical literature, an area fraught with difficulties in establishing causal effects. Section 3.1.iii looks at valuation, also a difficult area and one prone to double counting.

i) Transport improvement and quantity changes: Theory

What are the effects of reducing the costs of flows of services, goods, and people in and out of an area? The area has some productive assets that are immobile (land, existing buildings), others that are mobile with frictions (the labor force), and others that, over some time horizon, are quite mobile (firms and perhaps some other institutions).

The first approach to answering this question comes from trade theory (from Ricardo onwards). Better access will enable the area's export sector to expand and cause its import competing sector to contract, this in line with comparative advantage and thereby bringing economic gain. Later trade theories studied determinants of comparative advantage and highlighted that some factors of production would gain and others might lose. These are important insights. They are however only a first step to addressing the issue of within-country transport improvement, in part because they are based on perfect competition, inter-industry trade, and immobility of factors of production, which are all assumptions of limited relevance in this context.

[12] See Redding and Venables (2004) for international evidence on the cost of remoteness.

An alternative is to focus on factor mobility and, in particular, the location decisions of firms. The forces that shape these decisions can be put in four main categories: (i) production costs (the only force that matters in classical trade theory), determined by factor prices as well as technology and institutions: (ii) market access, depending on the size of the local market and on the ease of access to 'export' markets: (iii) product market competition, meaning the number of competitors and extent of competition faced by firms in local markets and export markets: and (iv) the 'business ecosystem', i.e. the number of other firms and networks that are complementary, rather than competitive with the firm's activities.

A transport improvement is a double-edged sword, and theory suggests that it can make a place either more or less attractive for firms. One obvious tension arises as transport both improves access to export markets and opens the local market to import competition. Another comes from the interplay between market access and production costs. A number of models find that reducing transport costs from a high level to an intermediate level is a source of divergence (one of the connected points may expand at the expense of others), while reducing transport costs further gives the opposite result, leading to convergence.[13] The ambiguity arises because market access matters a lot when transport costs are very high, but production cost differences become all important once 'globalization' has reduced transport costs to low levels. Box 3.1 explains the intuition behind this result.

The arguments of the preceding paragraph become more acute the greater is the mobility of people and firms. Since market size matters for firm location, a region that is losing population will become ever less attractive for firms; this may be offset by falling prices of immobile factors (land), but if land is only a small fraction of firms' and households' expenditures rent may not be able to fall far enough to stop the exodus.

Firms' location decisions depend also on the 'business ecosystem' – their access to other firms that are a source of intermediate inputs and markets for their output. This supports clusters of activity, and also means that there may be multiple equilibria; a cluster could equally well operate in one place or in some other. Its actual location may have been determined by historical circumstance not economic fundamentals, but once established it creates its own productivity advantage. This makes it even more difficult to predict the effect of transport improvement on location. A new place may appear to have the advantages of factor costs and market access needed to attract activity in some sector, but firms fail to move out of an existing center because they fear foregoing the benefits of the existing business ecosystem. And if change does occur it may be sudden, as new clusters develop and old ones unravel, a phenomenon seen clearly in the impact of international trade and globalization on many established centers of manufacturing

[13] See Fujita et al (1999), Helpman (1998), or Combes et al (2005).

production.[14] The possibility of multiple equilibria also points to the importance of complementary policies (including soft policies intended to create positive expectations about the future performance of a place) in achieving change.

----------Box 3.1: Divergence then convergence: Transport between two cities----------

Suppose that there are two cities, one larger than the other. A firm supplies customers in both cities and is considering where to base its production.

- If transport and communication costs are very high the firm will operate facilities in both cities; it cannot reach all its customers otherwise.
- If transport costs fall to a moderate level the firm can serve both cities from one facility, and it will choose to locate in the larger city where most of its customers are. As many firms do this it will tend to raise land prices (and possibly wages) in the larger city and lower them in the smaller one – causing further divergence of city size and income.
- Suppose now that transport/ communication costs fall further and become extremely low. The cost of reaching customers is now irrelevant compared to production costs, which include rent and wages. The firm will run its operations from whichever city is cheaper, this favoring the smaller city and – as many firms start to do this – reversing the previous divergence. The limit, when transport costs are zero, is trade theory's 'factor price equalization'.

While just an example, this serves to illustrate some of the fundamental difficulties encountered in predicting the effect of transport improvements on location of economic activity.

---

ii) Transport and economic outcomes: Empirics

Despite the central importance of the issue and much attention devoted to it in empirical work, our empirical knowledge is still extremely scarce and perhaps the precious little that we know is of limited applicability to new projects.

The first reason for this is the ambiguity discussed in the preceding sub-section. Even simple models suggest that effects can vary greatly according to details of the context, in view of which it is not surprising that an empirical consensus has not emerged. Real world contexts are inevitably much more complex than theory modeling, not least since the theory is often based on just two locations. This simplification is often needed to generate results and is sometimes justified when thinking for instance about broad patterns of economic integration between a center and a periphery. Unfortunately, this two-location simplification is far from innocuous in our context. In a two-location model, employment growth in one location can only come at the

[14] See Fujita et al. (1999) for development of these arguments at various spatial scales.

expense of the other. In reality, we need to think about locations treated by a transportation project and locations not directly treated but which may nonetheless be affected as firms and workers may relocate to the treated location.

A further difficulty regards the measurement of market access and changes in market access following a new transport infrastructure. A key insight from theory is that simple measures of market potential where, for instance, GDP or population in other regions is discounted by some measure of distance or transport cost, are inappropriate. The reason is that it is not just the size and cost of accessing neighboring markets that matters, but also the intensity of competition and prices that prevail on those neighboring markets. Being close to a large market is not necessarily a good thing if that market is served by many competitors. It may be better to be close to a smaller number of potential customers that are currently paying a high price. Hence, the price of goods should figure prominently in the calculation of market access. Unfortunately, we do not observe the price of goods directly except for very particular goods that are highly homogenous such as ready-made concrete. Although some solutions have been developed to measure market access indirectly (Head and Mayer, 2014), they tend to be highly demanding in terms of data. It is fair to say that this measurement problem has been a major empirical limitation. Rather than being able to explore the full causal chain that goes from a policy treatment (such as a road improvement) to its effects on trade costs, then to effects on market access and finally to final outcomes,[15] most of the empirical literature has taken a more reduced-form approach trying to estimate directly the treatment effect of new corridor (or improved corridor) infrastructure on final outcomes.[16]

Third, while ignoring the full chain of causal events was perhaps a reasonable shortcut, this has the unfortunate drawback of leading research to neglect a key lesson from theory. We should expect highly heterogeneous outcomes when lowering shipping costs across places.[17] Instead, most of the research has attempted to estimate average effects. The standard approach taken by research is a regression of a change in outcome such as local employment or productivity on a change in infrastructure (or sometimes an initial level of infrastructure, a valid approach when adjustments are slow).

[15] If we had reliable results on the effects of changes in market access, we could perhaps compute the market access effect of a project and then directly make predictions using these results.

[16] Following Donaldson (2010), there is a stream of more structural work that assesses the effects of corridor infrastructure. Most of that work however dampens the role of factor mobility since "production opportunities" are assumed either exogenous or randomly drawn from a distribution with a fixed support. Essentially, every location has a shot at producing something irrespective of its population. This type of assumption is certainly justified in situations where factors are poorly mobile like 19th century India but this drastically limits the type of cumulative-causation effects discussed above in a more modern context.

[17] A possible partial exception is Faber's (2014) analysis of the effects of the construction of the Chinese national highway system during the late 1990s and early 2000s for which he finds a negative effect on peripheral counties.

Even using this standard approach, further difficulties are encountered. One is simultaneity. Rational infrastructure planning requires placing infrastructure in areas where it has most impact. This obviously leads to a spurious correlation between the outcome of interest and the placement of infrastructure. Alternatively, political realities probably imply that public works take place in locations hit by negative shocks. To go around this problem, the literature has developed a number of empirical strategies that rely on the examination of "in-between" locations along corridors that were incidentally served, engineering/cost predictions, or old infrastructure/plans developed under very different circumstances to generate some quasi-random variation for the corridors that were developed.[18]

Another fundamental problem is distinguishing between displacement and net growth; if a treated place does better than an untreated one it could be because of pure displacement, with no net benefit. To be able to distinguish, the researcher needs three groups of units: a treatment group receiving direct access to the infrastructure, an indirectly treated group that may have suffered from displacement, and a control group of truly unaffected locations. Treated areas are relatively easy to define; for a new train line, they may be cities with new stations. Indirectly treated areas are harder to define; they may be cities that did not receive a new train station and are neighbors of treated cities, acknowledging that there are many shades of neighbors. The true difficulty is finding the valid control group. Assuming displacement effects decline with the distance to the treatment – which seems like a reasonable assumption although still an assumption – one would like a control group made of really remote cities. At the same time, these remote cities are likely to be different from the treated cities. Imagine for instance, a road improvement scheme in the island of Java in Indonesia that affects half the cities. Given the small size of the island, the other half is likely to be subject to displacements. For a control group, one will probably need to examine cities in the island of Sumatra. Whether Sumatran cities form an appropriate control group is unclear. They are likely to be subject to different dynamics. It is also perhaps hard to rule out the absence of displacements from Sumatran cities.

Notwithstanding these difficulties, there is now a large empirical literature on corridors – one too large to be reviewed comprehensively. The interested reader should refer to the reviews by Redding and Turner (2015) and Berg et al. (2017), the latter focused on developing countries. The literature points to four main findings so far. First, corridors tend to attract economic activity and at least some of this is driven by displacement from locations more remote relative to the infrastructure. Second, transport infrastructure also tends to promote the decentralization of economic activity within a corridor area away from the main centers. This dispersion of activity is nonetheless far from uniform along corridors. Third, corridors appear to promote various efficiency gains through higher productivity and less factor misallocation. Fourth, corridors affect several margins including the aggregate amount of economic activity in a

[18] See Redding and Turner (2015) for a thorough discussion of identification issues for transportation projects.

location, its distribution across sectors, its distributions across skills and functions (production vs. management for instance), and participation in external markets (labor markets in other locations or agricultural markets when moving away from subsistence farming). Given the methodological problems noted above, these findings should be viewed as tentative. They also reflect broad trends in the data but should not be expected to hold every time given the heterogeneity associated with relocations.

iii) Valuing transport improvements

The benefits of connectivity are widely touted and frequently double counted. To clarify the issues we start with simplest textbook thought experiment and then show how it needs to be generalized to capture developing country issues.

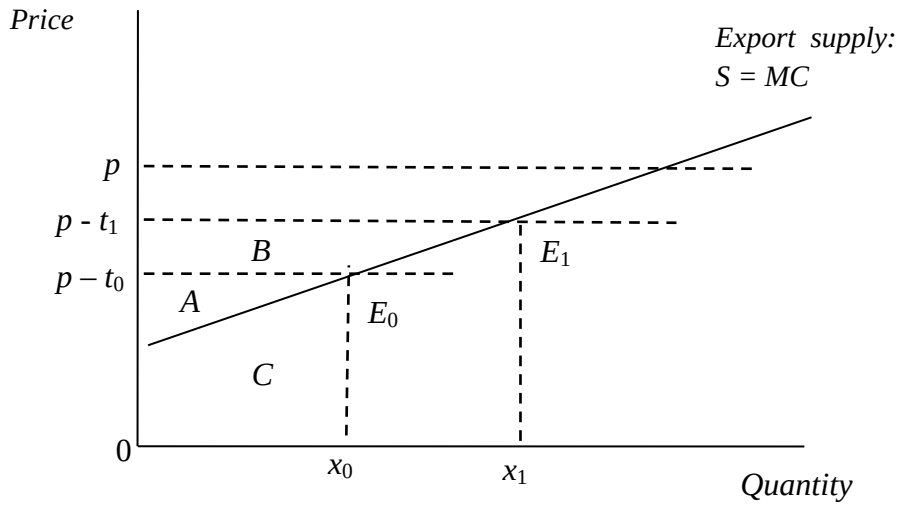
Consider a place that can 'export' some product, but only receives price [$p - t$] per unit, where p is the world price and t are the transport costs incurred in selling the product. For example, this could be an agricultural region, connected to its market by a road link. What is the value of improving the road link and reducing t? Figure 3.1 has price and unit values on the vertical axis and quantity sold on the horizontal. The demand curve is horizontal – we assume the world price of the product is fixed – and so too are unit receipts net of transport costs, given at two levels, t_0 and t_1, $t_0 > t_1$. The place under study has an export supply curve which is upwards sloping and is the marginal cost curve, S = MC. The initial equilibrium is at point E_0 and the new one, once trade costs are reduced, is at E_1, giving quantity change (additional exports) $x_1 - x_0$. The value of the trade cost reduction is area B. To see this, note that economic surplus in each situation is the triangle formed by the difference between total revenue (net of transport costs, initially rectangle $[p - t_0]x_0$ = A + C) and total production costs, the area under the MC curve (initially area C). The initial surplus is therefore area A, and the new surplus (with net price $p - t_1$ and quantity x_1) is A + B, giving gain of B.

This gives – in partial equilibrium framework – the gains from trade. Cost benefit analysts sometimes refer to this measure as the rule-of-half. If the change in unit transport costs is Δt, then area B is approximately:

Rule of half: $B \approx x_0\Delta t + (x_1 - x_0)\Delta t/2 = (x_1 + x_0)\Delta t\ /2$.

It follows from this, and from the figure, that the gain per unit quantity exported cannot be greater than Δt. Thus, the per unit gain from a \$1 reduction in the cost of exporting a unit of output is not more than \$1. The fundamental logic is that, if it were larger, then in the initial situation the exporter would have incurred the \$1 cost in order to reap the larger gain.

Figure 3.1: Valuing a transport improvement:



This argument is at the core of standard cost-benefit analysis and simply captures the direct or user- benefits (to existing and new users) of the transport system. It puts a tight cap on the value of transport improvements and is an application of the argument made in section 2 that, in a 'perfect' economy, induced changes in quantities are of no value. What happens outside this? We look at three main arguments.

Market failure and shadow pricing of inputs: The first argument stems directly from our discussion of market distortions in section 2.2. These imply that the private MC curve on which firms base decisions is not the same as the social MC curve. For example, if labor is being drawn from under-employment with low opportunity cost, then the social MC curve is lower, creating a further area of benefit. A way to capture this is to 'shadow price' labor, i.e. use the social opportunity cost of labor instead of the market wage in cost calculations.

Endogenous productivity and agglomeration: A further source of benefit or cost arises if a transport improvement and consequent private investment creates externalities. Some are negative (e.g. environmental) and others positive (less congestion) and many can be captured by shadow pricing (e.g. attaching a carbon price). In the spatial context a good deal of attention has been paid to the idea that transport improvement increases productivity by enabling the growth or deepening of clusters of activity. There are two distinct mechanisms through which this might occur. One is that, even given the location of economic activity, improved connectivity makes places 'closer' together in economic terms and thereby increases the productivity benefits of proximity and agglomeration. The other is that the transport improvement may trigger investment (quantity changes) which takes the form of firms moving into a cluster of activity, further raising productivity.

The second of these mechanisms turns on predicting quantity change which, as we saw above, is problematic. A transport improvement might well be associated with some clusters expanding and others contracting. The former is part of several countries' transport appraisal methodologies. Its implementation requires calculating the effect of transport on the effective density of each place, and then combining this with econometric estimates of agglomeration effects, as measured by the elasticity of productivity with respect to economic mass. We discuss this further in section 4.1, in the context of urban transport improvements.

Large investments and price change: Figure 3.1 and standard CBA is based on 'small' private sector firms that take the prices at which they trade – for inputs and outputs – as given. The development context seeks to look at situations where input prices – above all the wage – may be increased by development, and where some of the private sector investment decisions may be large relative to local markets. There is then the possibility that the value of transport improvements may be larger than suggested by standard theory.

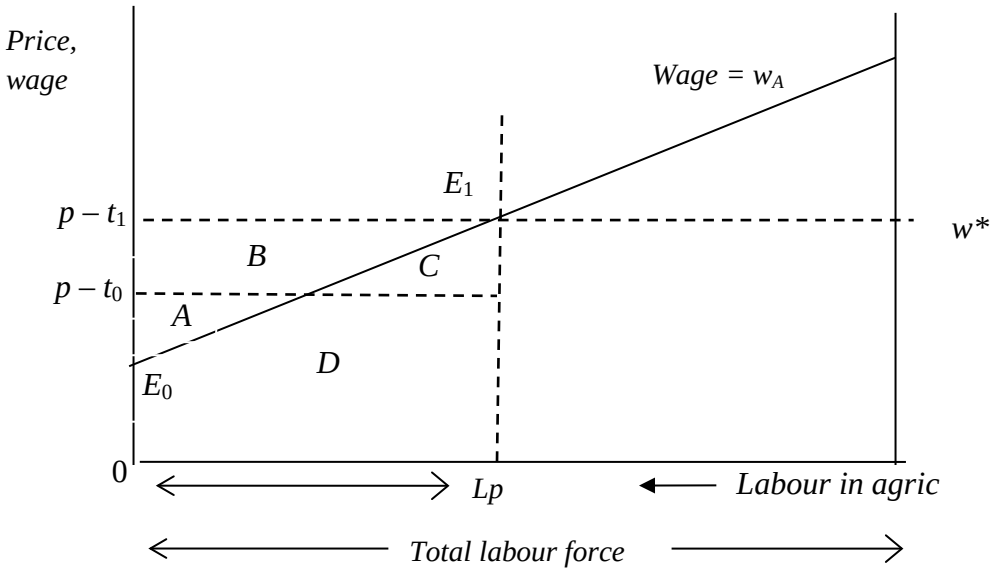
The argument is the following. Transport improvement might trigger private investment – a quantity response – that is large enough to raise input prices. In particular, there might be an increase in the prices of immobile factors of production. In some contexts, this is labor, or in others – such as an agricultural corridors or new business center in a city – the rental on land. Since these benefits do not accrue to the investor (unless, in the case of land, the investor acquired the land before announcing the investment), there will be a range of transport costs at which, while it is socially efficient to invest, the investment will not be undertaken. Lowering transport costs to the point at which the investment does occur then unlocks particularly large gains, potentially exceeding those suggested by the rule-of-half. The argument is made for a simple example in Box 3.2.

----------Box 3.2: Lumpy investments and price change----------

A region has total labor force measured by the length of the horizontal axis and labor is employed in either 'agriculture' (measured from the right axis) or 'processing' (measured from the left). The agricultural wage is the value marginal product, decreasing with labor employed, and given by the curve w_A; if agriculture is the only activity then E_0 is the equilibrium.

A processing firm is 'large' and can only operate at a scale where it employs (at least) L_p workers so has to pay wage w^*. Each worker produces one unit of output which is sold for $p - t$. At what value of t is it profitable for the firm to start up? At what value is it socially efficient to start? What are the gains from reducing t?

- It is profitable to enter if $t \leq t_1$, so the net price received, $p - t$, exceeds the wage per unit, w^*.
- Starting production at this point yields net social surplus, A+B > B: the value of processing produced is A+B+C+D whereas the value of agricultural output lost is C+D.
- It is socially efficient to enter if $t \leq t_0$, i.e. where A = C: the value of output gained is A+D and the value of output lost is C+D.



There are two messages:

- Since the firm is 'large' and therefore its actions change prices, it does not capture all the benefit of its investment – some is transmitted to higher wages of workers. This means that the firm will enter later (i.e. at lower transport costs) than is socially efficient.
- Threshold effect: Some transport improvements have no effect as they are not large enough to cross the threshold at which production starts. But a small improvement that crosses this threshold causes a welfare gain that is much larger than that predicted by the rule-of-half.

3.2. Special Economic Zones

The principle objective of SEZs is to attract investment and create jobs.[19] Ideally, this is investment in internationally footloose activities that would not have otherwise come to the country, and hence jobs that are additional not just displaced. Successful SEZs will link to the rest of the economy and stimulate growth more widely, with the ultimate objective of convergence between well-performing zones and the rest of the country. Further benefits might flow from supply of foreign exchange (in the investment itself and through the sale of output) and from the fact that SEZs can be a way for governments to experiment and find out what policies work.

SEZs employ a range of hard and soft policies in a well-defined geographical area or areas. Among hard policies employed are:

- Liberal treatment of imports and exemptions from customs duties, particularly (but not only) in the Export Processing Zones and Free Trade Zones that were precursors to current SEZs
- Tax incentives, particularly holidays from corporate income taxes
- Infrastructure provision, including electric power, transport, water and sanitation
- Distinct regulatory regimes, often involving laxer labor regulations, restrictions on union activity, and different land tenure systems
- Provision of large parcels of land, often with industrial sheds built in advance of occupation.

Soft policies include:

- Management of SEZs that seeks out and works closely and effectively with private sector investors
- Promotion of particular sectors
- Effective implementation and management
- Labor training.

The economic case for pursuing these policies in tightly defined geographical areas rests on two arguments. [20] One is 'first-best' and based on the economic efficiency gains derived from spatial concentration in the provision of infrastructure and development of clusters, as discussed

[19] We define SEZ broadly to encompass free-trade zones, export-processing zones, or any special district with favorable fiscal or institutional treatment.

[20] We only discuss economically motivated SEZs. There are obviously cases of SEZs that were set up to fulfill political objectives. For instance, the Tierra del Fuego province at the extreme southern tip of Argentina was designated as an SEZ to increase the local population in that sparsely populated region by attracting manufacturing activity, the latter being seemingly only an intermediate objective.

above. The other is 'second-best', based on the presence of institutional and financial constraints that create economic inefficiencies. These could, in principle, be removed economywide. But in practice this is infeasible, partly because of their fiscal cost (tax and customs regulation, infrastructure) and partly because of the political obstacles that would be encountered (for example in acquiring land and implementing regulatory reform).[21]

i) Direct quantity effects: Attracting investment

As outlined in section 2.2 above, success in attracting investment requires that a wide range of conditions be met, covering geography, policy, and the business ecosystem. There is a weakest link problem, and many SEZs have failed because key elements of the package are absent.

First, SEZs need to be located in places consistent with their objectives and long-run economic viability. If they are export oriented (or import dependent) they need to have good access to port infrastructure. In countries where even well-located regions have difficulty attracting investment, SEZs in backwards or remote regions are unlikely to succeed. Economic scale of the SEZ itself (to reap scale and agglomeration economies) and of the area where it is located (to provide a local labor market and depth of local firms) are important factors. The importance of these factors is confirmed in Farole's (2011) study of African SEZs. However, finding a 'good location' is necessary but not sufficient for success.

The success of an SEZ will also be determined by its comparative advantage, so unrealistic sectoral selection will lead to failure. Bangladesh initially sought to establish SEZs in high technology sectors and only after the decision was made to switch to labor-intensive garment manufacture did its SEZs take off. The fact that Bangladesh and other countries now occupy this product space has implications for the sort of sectors in which African countries could hope to be successful. In some sectors regional comparative advantage may be relevant rather than international comparative advantage.

There is considerable evidence that tax incentives alone are insufficient for success. Assessing the marginal impact of one policy is difficult given the complementarities between them and the country context (policies in place outside the SEZ). Nevertheless, Farole (2011) looks at data across 77 countries and finds that infrastructure and trade facilitation have a significant positive impact while tax and other financial incentives are much less important. This is consistent with recent Indian experience, where SEZs concentrated on tax and regulatory breaks while neglecting infrastructure provision. Khandelwal and Teachout (2016) compare firms inside and outside SEZs in India and Bangladesh. In Bangladesh, firms in SEZs experience much shorter

[21] This may come at a cost. As shown by Grant (2017), special economic zones in the United States are used to limit the opposition to protectionist policies that affect infra-marginal firms. In turn, this may allow policy distortions to be worse than they would otherwise be. In a development context, SEZs may allow strong unions to keep their lock on the labor market outside the SEZs, which might otherwise be challenged more seriously.

delays in getting import licenses and clearance, construction permits, and connections to water and power. In India most of these variables were worse for firms inside SEZs compared to outside (and worse than for Bangladeshi SEZ firms). Indian outcomes have been poor, with far fewer jobs created in SEZs than anticipated. The Arab Republic of Egypt provides another example.[22] For unclear reasons, it sought to restrict manufacturing activity to SEZs, many of them located in remote peripheral regions. This created upwards pressure on land prices in SEZs, to which government reacted by deciding to allocate land without a market, giving it for free to a lucky few. This is arguably not the best way to foster industrial development in a country.

In some situations, SEZs may be able to remedy one key bottleneck for development. For instance, the main advantage of Bangladesh's SEZs seems to be their ability to offer well-serviced land to manufacturing investors.[23] To take another example, the Government of Afghanistan is considering retrofitting former U.S. military bases as SEZs. Offering a more secure environment is arguably a crucial issue in this country. If the objective is to attract internationally mobile firms, then the reference level is meeting international standards, not just surpassing local standards.

Effective implementation of policy matters. This requires action that is coordinated across functions (tax, land, infrastructure), so it requires that the organization running the SEZ is empowered to deliver these functions. There must also be credible commitment to policy for many years ahead. Taken together, these considerations mean that commitment is needed from the highest level of government. At the same time the SEZ authority needs to be responsive to the concerns of firms in the zone.

ii) Indirect quantity effects: Links and growth

A successful SEZ will have an internal dynamic of spillovers between firms, agglomeration, and productivity growth. This will have a horizontal element, with a large number of firms in the same sector building up thick labor markets and other agglomeration economies. And a vertical element, with co-location of input suppliers and the growth of forward and backward linkages. This process encounters the first-mover or "all-at-once" coordination problem discussed above – it is hard to start a cluster. Involvement of one or several large firms is one route to kick-start this, as with the multinational electronics companies (including AMD, Fairchild semiconductor, Intel) initially attracted to Penang Malaysia, or Philips-van-Heusen's project in Hawassa, Ethiopia (see Box 3.3). Attracting such companies requires intense 'soft policy' from government, working closely with the companies and committing to deliver international standards.

[22] World Bank staff, personal communication.

[23] World Bank staff, personal communication.

Links from the SEZ to the local economy include development of skills in the local labor market; expanding the technological capabilities of local firms; increasing use of local firms as suppliers and as customers; and entrepreneurial spinoffs from firms in the zone. Successful SEZs have seen an increasing fraction of activity in the SEZ being undertaken by local firms, this sometimes occurring as part of a maturing and upgrading process. In Mauritius the SEZ upgraded from low-value textiles to higher value and more skill intensive products (off-shoring low-value production to the SEZ in Madagascar). In Malaysia the Penang SEZ focused from the start on electronics, but upgraded from basic assembly to more advanced and skill intensive goods. Both these sectoral transitions were accompanied by a transition towards locally owned firms.

The role of government in this process is important and needs to be based on recognition that there are mutual benefits – for firms in the SEZ and for the local economy – from developing these spillovers. Thus, rigid domestic content requirements are likely to be viewed as a cost to firms in the SEZ and may transfer little learning to firms outside. But working to bring local firms up to the level where they are chosen suppliers is of mutual benefit. The knowledge transfer is also of value to government itself, as SEZs can provide a vehicle for learning about what makes an effective business environment. China explicitly used SEZs as vehicles for policy experimentation.

iii) The value of SEZs

The costs and benefits of an SEZ depend on the quantity response elicited and on displacement – the extent to which investments and jobs created are additional to those that would have occurred absent the policy. The value of jobs created depends on the state of the local labor market and the alternative sources of employment, as we saw in section 2. Linkages to the local economy should be included to derive the net number of jobs in the economy relative to a situation without the SEZ policy.

Benefits accrue directly through (net) job creation in the SEZ and also through potential impacts on the wider economy. One mechanism is sheer scale: in Bangladesh and Mauritius the scale of job creation (in the SEZ, in suppliers, and via spending from wages) surely raised incomes not just of those employed in the SEZ, but through tightening the labor market throughout the country.[24] Other mechanisms operate through raising skills and capabilities of workers and firms both inside and outside the zone, and through the consequent dynamics of productivity growth and increasing competitiveness in international markets. The gains are potentially substantial, but as suggested above achieving them requires meeting substantially all of a large set of conditions.

[24] Technically, this delivers a 'terms-of-trade' gain, arising as higher wages are earned in export activities.

----------Box 3.3: Hawassa Industrial Park----------

The Hawassa Industrial Park (HIP) was inaugurated in July 2016 and is planned to create 60,000 direct jobs on $1bn export sales from the park. Its distinctive features have created optimism about prospects for success.

The project is based on close collaboration between Philips-van-Heusen (PVH) and the government of Ethiopia through its investment agency and the HIP agency. Areas of mutual interest were established – including building world-class factories with the highest environmental standards – and the government demonstrated commitment by raising capital and building facilities fast. PVH brings the advantage of scale (it is the world's second largest apparel company) which mitigates the first-mover problem. It plans to develop its own production ***and*** a local supply chain. Its purchasing power means that (as of 2017) some 18 foreign and 5 domestic supplier companies have committed to follow it to HIP and further foreign and local suppliers are planned. Vertical integration extends to textile mills and will continue up through the value chain, with100k ha of land purchased for cotton growing.

What made PVH think that Ethiopia was the place where such a project could succeed? Access to international markets was essential, including the EBA agreement, the 10-year extension of AGOA, and streamlined customs procedures. The government is extending the new Djibouti-Addis Ababa rail line to reach Hawassa, and the town also lies on the improving Addis-Ababa to Kenya highway. Electric power is a major cost in the garment business, and Ethiopia's recently expanded hydro-capacity enables the country to undercut other countries in the region both in price and reliability. Hawassa offers a large pool of local labor, and employee selection and training programs are being run by PVH jointly with the government and donors. Tax holidays are offered but are described as 'icing on the cake' rather than key elements of the decision.

A key element of the project is the role of two large and committed players. Government commitment was crucial to PVH's decision and was manifest from the early stages of formulating objectives, demonstrably followed through with implementation, and thereby creating expectations of future cooperation. On the other side, PVH is a 'large developer', able to secure the movement of supplier firms and overcome the first-mover problem.

This box draws on Mihretu and Llobet (2017) and conversation with Martin Green.

Finally, the costs of the policy depend on the set of instruments used. Tax breaks appear expensive but have to be compared with revenue that would have been earned absent the SEZ; compared to this counterfactual they are costly only if they divert tax-paying firms into the zone, rather than create new investment in the zone. Infrastructure investment is riskier, since costs are incurred at early stages of development while benefits depend on the success of the SEZ. Regulatory innovation and soft policy is a low-cost policy from which the government learns, even in the event of failure.

3.3. Lagging regions

Economic development is inherently spatially uneven. Some regions lead development and others lag; indeed, it would be remarkable if all regions were to develop at the same rate. In the absence of policy, what happens to regions that are lagging? They often catch up as activity spreads out of an economic core – or spreads inland from the coast – as is documented in WDR (2009). But in some cases, they fail to converge in this way. Many countries have regions with chronic problems, such as Brazil's Nordeste, China's Xinjiang region or India's state of Bihar, to name just a few.

What are the economic issues that prevent catch-up from being the usual outcome? There are two main factors at play. First, as discussed in section 2, factor mobility and intra-country institutional factors limit the magnitude of interregional factor price differentials, implying that the principal mechanisms of comparative advantage do not operate. Interregional price differentials will open on prices of immobile factors of production (land and housing) but much less so on labor and capital and may be jointly insufficient to attract new investment. Adjustment then takes the form of labor moving out of lagging areas. Notice that there is no market failure directly associated with this mechanism. In these circumstances, it is efficient to let the region contract.

The other economic issue preventing catch-up is the propensity of activity to cluster and the difficulties of starting new centers of activity. As we saw in section 2, the productivity gains from clustering are a positive externality, although coordination failure can mean that economic centers become too large. These market failures mean that the equilibrium is inefficient, despite net gains from clustering. However, while these considerations suggest a case for decentralizing activity from a congested city to a secondary or satellite-city, they do not make the case for decentralizing to a remote region.

Given this, what is the case for policy intervention? It typically rests on one or more of the following arguments. First, the negative impact on people left behind. This can be acute in areas of absolute decline where out-migration of the young (and possibly more skilled) leads to demographic imbalance and severe social deprivation. More generally, persistent inequalities raise issues of spatial equity. Beyond these normative considerations, we note that profound spatial disparities may lead to distributional conflicts and other political ailments. Second, the argument is made in terms of excess (or over-rapid) expansion of booming areas, leading to congestion and pressure on housing and other assets. However, as suggested above, better policy response to this is likely to be policy within the booming region, rather than in some other region. Third, some particular market failures may be holding a region back. Although these market failures may occur everywhere, they be may be particularly harmful in lagging regions. A stronger version of this argument is that lagging regions may be stuck in a poverty trap.

Any case for policy intervention ultimately depends (one hopes) on a belief that policy can be effective, and it is important to be clear about what success means in this context. For purposes of this paper, two criteria matter. One is based on the direct effects of policy. Is policy cost-effective in delivering a narrow set of outcomes, such as improved public services or journey times? The other is the 'indirect effects', which boil down to attracting investment and creating jobs such that the region has long-run economic viability which it would not have otherwise been expected to achieve.[25]

i) Policies for lagging regions

Interventions are often part of ambitious multi-instrument programs that include (a) transport investments to improve connections within lagging regions and between lagging and more prosperous regions; (b) fiscal incentives and various direct service provisions; and (c) a package of measures that aim to foster skills, enterprise development, and innovation in specific parts of a country. We group these measures under the heading of local economic development (LED) policy.[26]

We already discussed transport connections, fiscal incentives, and local service provisions. We do not repeat these discussions here but note that the effects of transport connections are often ambiguous while, as with SEZs, geography and the scale of these zones matter. Instead, we focus our attention in this section on two key aspects of policies for lagging regions. The first is the use of multiple and potentially complementary instruments in ambitious regional development programs that attempt to make a large difference to economic outcomes reasonably fast. These policies are often referred to as "big push" policies. Our second, narrower focus, will be on LED policies.

Before going further, we make several important notes of caution. The first is that knowledge about what works is weak. Second, prospects depend on many characteristics of the region that are not amenable to control by economic policy measures. Third, for treatment to be effective it needs to be geographically selective. It follows that relative decline of some areas will, in many countries, be an inevitable part of national economic development.

[25] By 'viability' we mean ability to pay wages comparable with those in the other regions, while attracting investment and supporting employment. All regions need to have a source of 'foreign exchange' if they are to finance necessary 'imports'. This is clear at the national level and also holds sub-nationally, referring in the sub-national context to earnings or other revenue received from the rest of the country rather than (necessarily) the rest of the world. Thus, a viable region has to be competitive selling goods and services outside the region. If not, it has to be dependent on transfers from central government or other sources – an outcome that we do not deem successful.

[26] Countries also often use public employment to foster regional development. More public employment may help increase local capabilities, but this seems doubtful. Local public servants will spend a large share of their income locally. As such, this type of policy is essentially a form of transfer. We also note the public employment may displace employment in the private sector (Faggio and Overman, 2014).

ii) Complementary policies and the big push

The usual justification for large comprehensive packages lies in either strong complementarities between policy instruments or the existence of local poverty traps (and often both). For instance, providing a transport link to a peripheral region in a developing country may not generate positive effects if local producers are unable to benefit from better access to new markets. This transport improvement thus needs to be supplemented by some capacity building for local firms. More generally, in economies plagued by numerous market failures and inefficiencies, fixing a single problem does not, in general, offer a guarantee of improvements. This is the classic second-best argument of Lipsey and Lancaster (1956). To take a simple example, tolling a major road to avoid congestion may push traffic to secondary roads and increase the number of accidents. The social loss caused by more accidents may dominate the congestion gain, resulting in a net social loss.

The main challenge with policy packages that hope to build on complementarities between instruments is that they require an extremely detailed understanding of what the frictions and market failures are and how they interact. In some cases, these complementarities seem obvious. For instance, both electricity and market access are arguably necessary for an export-processing zone to be successful. Other interactions are much less well understood, including those that take place between hard infrastructure such as transport improvements and softer interventions such as LED policies. While being deep in the second-best offers no guarantee that remedying one market failure will generate an improvement, neither does it guarantee that fixing two market failures will do better.

Turning to multiple equilibria, we note that being able to move from a low to a high equilibrium sounds extremely attractive since a temporary intervention may be able to make permanent change. Even more attractive, this move could potentially be achieved at relatively low cost even when the high equilibrium is much more desirable than the low equilibrium. This is the logic that some policy makers appear to have in mind when they combine some local capability developments, support for higher education, and relocation incentives to technology firms in the hope of creating a “transformative” high tech cluster.

This type of intervention raises two issues. First, the knowledge needed to design them far exceeds what is currently known. As we argue below, extant research still struggles to offer solid conclusions regarding the existence of multiple equilibria in regional development and our ability to move from one equilibrium to another. Second, the notion of multiple equilibria is more complicated than it seems. We show this in box 3.4. In some cases, multiple equilibria are zero, if not negative sum games.

----------Box: 3.4: Single vs. multiple equilibria in regional development ----------

Although they all involve some form of increasing returns, it is useful to distinguish between different forms of multiple equilibria. Some are not explicitly spatial, as in the literature on poverty traps. While there are many ways to justify the existence of poverty traps, a first classic motivation is the existence of a low human capital and low productivity equilibrium, which can persist when a region cannot finance its initial growth in human capital (Azariadis and Drazen, 1990). Another classic example involves the existence of a persistent agrarian equilibrium while the industrial sector operates under increasing returns but requires enough local demand to be viable which is not the case in an agrarian economy (Murphy, Shleifer, and Vishny, 1989).

In the spatial context, we distinguish two types. Think of an economy with two regions. The first type is where economic activity is concentrated in one region or the other, but which is not determined ex ante. While the Silicon Valley emerged around Palo Alto in California, New Jersey with its high concentration of human capital was also a strong candidate. Naturally, the lagging region wants to change the equilibrium it faces to become the new winning region. However, any policy to that effect is of no social value, unless the concentration of economic activity initially occurred in the "wrong region".

A second type arises when concentration in one region and an even distribution across regions are both in equilibrium (Krugman, 1991a). Concentration of economic activity may be in equilibrium as no firm wants to leave the leading region where they benefit from agglomeration effects, and serving the leading region from the peripheral region would be too costly. Under the same fundamentals, an even distribution of economic activity may also be in equilibrium because the benefits for firms of increased concentration in one region would not compensate the higher cost of serving the equally large market in the other region. Which configuration is more efficient depends on the fine details of the exact model at hand (Baldwin et al., 2005). More generally, various configurations with various degrees of asymmetry may occur in equilibrium and there is no general result about which configuration is more desirable.

A key question with multiple equilibria is how to move from one equilibrium to the other. In absence of friction, a shock is needed to take the economy away from the basin of attraction of the low equilibrium into the basin of attraction of the high equilibrium. With transitional frictions such as the cost of rural-urban migration or the inability of an industrial sector to absorb new workers fast, the situation is more complicated. The frictions may be large enough to trap the economy in the low equilibrium. When frictions are less, expectations become fundamental (Krugman, 1991b, Matsuyama, 1991). This opens a role for the government to coordinate expectations to avoid the same type of coordination failures as with urban development (see below).

----

iii) Evidence on big push policies and multiple equilibria: Quantity and price effects

While countries or groups of countries in Europe often tout big-push type of initiatives involving multiple interventions with the stated objective of setting a lagging region on a "different path", the reality on the ground is often more modest. For instance, there are various European Union funds that aim to promote convergence among European regions, their competitiveness, and their

cooperation.[27] These funds support an extremely wide variety of interventions from innovation and job creation to labor market inclusion, training, and infrastructure. This said, the overall budget for these funds represents only about 0.35% of European GDP and is smeared broadly, as a majority of European regions can qualify for one type of funding or another.

Examples of big-push interventions from the developing world abound. For instance, the Upper-Egypt Local Development Program targets two provinces with a variety of interventions involving infrastructure development, the construction of special economic zones, and a variety of support programs for both private and public sectors. Argentina has recently launched its Plan Belgrano for its ten (lagging) northern provinces with a mix of infrastructure, private sector development measures, housing construction, and increased childcare provision. While ambitious in its purpose, the investments proposed by the Belgrano plan represent only about 0.25% of the country's GDP annually for a period of 10 years.

This type of wide-ranging intervention is extremely hard to assess empirically. A first possibility is to try to assess every single instrument individually, but this would be daunting. This would also be contradictory since the rationale for these instruments relies precisely on the complementarities between them. A second possibility is to take a more aggregate approach and only examine final outcomes such as overall employment or GDP per capita in the treated regions relative to the untreated. The problem is then that these aggregated outcomes may have been affected by other aggregate changes in the economy unrelated to the policies at hand. Displacement will also affect "untreated" regions. Regardless of the approach taken, the difficulties of the evaluation are further compounded by the wide geographical dispersion of the investments, the modest amounts being invested, and the ongoing nature of many of these policies with few abrupt changes. This is arguably why there is little research on these projects and why most of extant research is struggling to provide solid conclusions (Neumark and Simpson, 2015).

The main exception is the wide-ranging evaluation of the Tennessee Valley Authority (TVA) program by Kline and Moretti (2013b). Like the interventions described above, the TVA had several components, mainly energy generation (through the construction of numerous dams), transport (through the development of roads and canals), and education (through the construction of new schools). For evaluation purposes, the TVA has the advantage of being well circumscribed in time (starting in 1933 but with the bulk of the investments made in the 1940s and 1950s) and geographically (163 counties across four states in the Appalachia region). While small relative to the scale of the U.S. economy, the transfers were substantial for the treated counties, up to 10% of local incomes at the beginning of the 1950s.

[27] There are five main funds: the European Regional Development Fund, the European Social Fund, the Cohesion Fund, the European Agricultural Fund for Rural Development, and the European Maritime and Fisheries Fund. There is also a European Solidarity Fund to provide emergency support in case of major disasters.

Kline and Moretti (2013b) examine the evolution of TVA counties vs. control counties over 1940-1960 when the program was active and over 1960-1980 after transfers had ended. Relative to their control group, TVA counties enjoyed higher growth in manufacturing employment of 5-6% per decade and lower growth in agricultural employment. Much of the growth in manufacturing occurred during the treatment period while the decline in agricultural employment occurred after 1960 and the end of transfers. While median family incomes in TVA counties increased by about 2.5% per decade relative to control counties, manufacturing wages did not increase. This suggests that the main effect of the program in the treated counties was to industrialize them by shifting labor away from agriculture into manufacturing where wages were higher.

To value these sizeable changes in quantity, Kline and Moretti (2013b) allow productivity to increase in TVA counties either directly through infrastructure investments or indirectly through agglomeration effects. As we argue below, for agglomeration effects to be of value, they need to be stronger in TVA counties than in the rest of the country. Kline and Moretti (2013b) do not find any evidence to that effect. However, they find that the direct productivity benefits exceed the federal transfers by about 30%, which puts the program at the margin of profitability given that public funds come at a cost.

An important caveat when doing an aggregate evaluation of this type is that displacement effects cannot be tracked directly.[28] Just as with corridor projects, one would need to observe "control" counties in the United States had the TVA program not taken place, which is of course impossible. With this problem in mind, Kline and Moretti (2013b) impose some theoretical structure to interpret their findings and indirectly estimate displacement effects.[29] A key challenge for future research will be to assess the sensitivity of this type of analysis to the details of the model at hand.

Importantly, Kline and Moretti (2013b) show that the effects on manufacturing employment of the TVA program can still be observed in 2000, about 40 years after federal transfers stopped. This suggests a change of equilibrium where a permanently higher manufacturing specialization can be sustained through agglomeration effects.[30]

[28] Over a long period of time like the one considered by Kline and Moretti (2013b), displacements can take the form of plants that physically relocate to the treated area. These perhaps can be tracked. Displacement may also take the form of plants that opened in treated counties during the treatment when they would have otherwise opened elsewhere in absence of the program. Tracking such displacements is extremely difficult.

[29] More specifically, the model used by Kline and Moretti (2013b) does not consider any friction in the adjustment process. These frictions are nonetheless at the heart of models of equilibrium selection (Krugman, 1991b, Matsuyama 1991).

[30] There is a complication here. As we show below, the existence of agglomeration effects can in theory allow for sustained economic activity in a location where previously there was none or very little. This is however subject to the condition that "intrinsic" productivity in this location is high enough so that in combination with agglomeration

At the heart of big-push types of policies is the notion that a local economy may be shifted from a less desirable equilibrium to a more desirable equilibrium. While direct evidence about big push interventions is limited, some insights are gained from the fast-developing literature that focuses more directly on multiple equilibria and assesses resilience vs. change in urban and regional contexts, following the work of Davis and Weinstein (2002, 2008). Inspired by the predictions of multiple equilibria in models of regional development (e.g., Krugman, 1991a) and the models of history vs. expectations described above (Matsuyama, 1991, Krugman, 1991b), this literature examines the effects of shocks on various local outcomes, from aggregate population, to patterns of specialization, and to house and land prices at a fine spatial scale.

While it is hard to know how relevant and representative the cases examined by the literature are for our purpose, we note that a majority of studies provide evidence of strong local resilience despite sometimes extreme shocks. For instance, Davis and Weinstein (2002, 2008) report that Japanese cities quickly re-converged to their pre-war relative population and even to their pre-war patterns of economic specialization, despite a good proportion of these cities having suffered extensive bombing during World War II. Bleakley and Lin (2011) document that portage stations along the Fall Line on the Southern part of the U.S. Eastern Seaboard continued to prosper long after portage disappeared, following first the construction of locks and then the decline of canals.

Although many of the findings are suggestive of a strong resilience of local equilibria, there are nonetheless documented examples consistent with a change in equilibrium. Redding, Sturm, and Wolf (2011) examine an extreme case: the location of airport hubs in Germany before and after the separation of the country after World War II. Berlin essentially lost its position as the main hub of German aviation, to the benefit of Frankfurt, and did not regain it after reunification. Siodla (2015) documents a strong discontinuity in patterns of land use and development for the parts of San Francisco that burned down following the major 1906 earthquake. Hornbeck and Keniston (2017) report similar findings following the great Boston fire of 1872. Importantly, Hornbeck and Keniston also document the mechanisms through which changes occurred, most crucially the redevelopment of buildings in areas that burned down, perhaps leading to positive neighborhood externalities.

Put together, this literature appears supportive of the models originally proposed by Krugman (1991b) and Matsuyama (1991). We may observe a change in equilibrium when some externalities are at play and when the frictions involved are weak. The existing stock of buildings

---

effects, achieved productivity is high enough to make this location economically viable. What Kline and Moretti (2013b) show is a "weak form" of multiple equilibria where an initial investment raises productivity and is compounded by agglomeration effects. Then, this investment can be sustained using local resources (which was the case with the TVA after the federal transfers stopped in the late 1950s). A "strong form" of multiple equilibria would only involve agglomeration effects. Kline and Moretti (2013b) cannot directly disentangle between the productivity effect of the initial investments and those of agglomeration.

seems to play a particularly important frictional role in practice.[31] Changes between equilibria appear to occur more easily at small spatial scales, districts within cities instead of regions within a country. This is perhaps consistent with the greater importance of agglomeration externalities and the greater ease of coordinating expectations at small spatial scales.

In conclusion, the empirical evidence invites caution with respect to big-push type of initiatives. First, existing equilibria often appear extremely resilient so that quantity changes are hard to generate through a change of equilibrium.[32] Second, the absence of "non-linearities" in agglomeration effects is consistent with quantity changes being of low, perhaps zero, values.

iv) Local economic development policies: Some possible justifications

The usual justification for LED interventions and support to innovation is that some regions have insufficient "capacities" because of lack of investment. The objective of the policy is then to foster capacity development either directly through the provision of skills and advice to firms or indirectly through fiscal and monetary incentives.[33] The key justification for investment subsidies or direct investment by the government lies in the existence of a wedge between private and social returns to investments. For instance, workers may underinvest in skills and education because of human capital externalities. In developing countries, credit constraints may further impair investments in education and skills. A lack of information may also be invoked.[34] For innovation, the justification is even more commonly accepted given the public good nature of new knowledge.[35]

[31] While Krugman (1991b) and Matsuyama (1991) assume fairly mechanical frictions associated with the movement from one location to another or from one sector to another, Glaeser and Gyourko (2005) highlight a further mechanism associated with the durability of housing. As residents start moving out of a location, the price of immobile assets falls, which dampens the incentive to move of the remaining residents.

[32] From an evolutionary perspective, this conclusion is not too surprising. "Fragile equilibria" may have disappeared following earlier shocks.

[33] Some policies seek to improve the capabilities of local governments to achieve better production and delivery of public goods. The justification for these policies is self-evident.

[34] A lack of information normally calls for no more than delivering this information to the economic agents who need to know about it. In an experiment in the Dominican Republic, Jensen (2010) finds that 8th graders who are told about returns to education during a brief session in class stay for about three months longer in school relative to similar pupils who are not told. While this type of quick and simple intervention appears extremely cost-effective, it may not be enough to ensure an efficient investment in education by youngsters given that this information may be difficult to transmit credibly, that decisions to acquire education are arguably subject to various behavioral biases, and that these decisions are subject to various other frictions such as the need to convince parents as well.

[35] For brevity, we do not provide a detailed discussion here. There are nonetheless important subtleties. Let us just mention one. Firms may underinvest in innovative activity because they are generally unable to capture the social benefits that arise from increasing the stock of knowledge and thus fostering further innovation. At the same time, competing firms may overinvest in innovation relative to what would be socially desirable through a "market stealing" effect. A quality improvement or a cost reduction may lead to magnified increases in profits leading firms to overinvest in innovation. It is a priori unclear whether over- or underinvestment prevails. See Aghion and Howitt (2008) for further discussions of innovation policies.

A clear limitation to this type of policy is that underinvestment in skills and various forms of capital may be a symptom of a deeper problem, rather than the problem itself. Economic agents choose to invest little when private returns are low, but these returns may be low for many different reasons. In some cases (such as human capital externalities) private returns may be low while social returns are high. In other situations (such as when jobs are allocated along group membership) both private and social returns are low. In these cases, subsidizing investment in skills is wasteful. The problem is elsewhere. The general point is that the appropriate subsidy requires diagnosing what the market failure is, and thereby knowing the wedge between private and social returns.

Diagnosis of market failures may be hard, as they often arise in related markets. An example is enterprise development and entrepreneurship decisions. Consider a simplified situation where an efficient allocation of talent requires that the most talented people become entrepreneurs. However, entrepreneurship is highly risky and there is no insurance market for entrepreneurial risk. This leads to too few entrepreneurs and misallocation where highly talented but also highly risk-averse individuals may choose to remain in occupations where their contribution to social welfare is low. In developing countries, other frictions leading to potentially severe misallocation loom large. Credit constraints are one of them. Cultural and religious traditions that reserve some occupations to some particular groups and bar others are another source of frictions.

While a strong case can be made for subsidies to skill acquisition, enterprise development, and innovation, complications arise when thinking about such policies at the subnational level. First, why should those policies be conducted locally and / or for only a part of a country? The market failures we just described are general. While it can be argued that specific contexts may call for specific responses, this does not justify a specific focus on lagging regions.

It could be that underinvestment in capacity is worse in lagging regions. But this needs to be justified and explained. We may expect social returns to investment in skills to be larger than the private returns everywhere in a country. It is unclear why this wedge between private and social returns should be larger where the private returns are the lowest. There are situations where this may be the case. For instance, similar matching frictions on the labor market will disproportionately affect regions where productivity is lower. In turn, this may justify labor market policies specifically targeted to less productive regions (Kline and Moretti, 2013a). Credit constraints may affect entrepreneurs and enterprise development disproportionately in poorer regions, but there are also situations where the opposite may hold, and more subsidies may be needed in richer regions. For instance, subsidies to innovation will arguably be much more effective when distributed to firms that are at the technological frontier than when given to firms that are well-within that frontier. The broader question behind this discussion is to what extent different policies are needed in different subnational contexts relative to national policies for which one size should fit all. A full treatment of this question goes much beyond the scope of

our discussion here. We note however that the greater spatial disparities in much of the developing world (Gennaioli et al., 2013) may warrant more policy differentiation.

v) Human capital and enterprise development in lagging regions: Quantity and price effects

Research is strongly suggestive that more human capital has large local productivity effects that can be tracked through higher wages and higher population growth. See box 3.5 for further details and references. Most of this work has been done in developed economies, but research in developing countries is suggestive of similar effects possibly with larger magnitudes for China, India, and much of Latin America (Duranton 2016, Chauvin et al., 2017, Ferreyra, 2017). For LED policies, these findings are good news since relatively small changes in the composition of human capital are potentially at the source of large changes in wages and employment.

This is only half of the story however. The other half regards the ability of LED policies to change the skill-composition of the workforce. Here the news is much less optimistic. There is a large literature looking at human capital policies at the country level which is beyond our scope. There is also a large literature that evaluates a wide variety of training programs, primarily in developed countries but also sometimes in developing countries. While there are a variety of results, overall the returns to training and other localized human capital interventions are found to be low and even often negative some of the time (Heckman et al., 1999, Card et al., 2010).[36] In developed countries, there is also some evidence that returns to training may be lower in high unemployment areas (Hyman, 2017).

A particular concern when focusing on human capital policies conducted locally is the possibility for educated or trained workers to migrate away. While outward mobility is likely to attenuate greatly the quantity effects of local human capital policies locally, there is clearly social value in raising the human capital of workers and their ability to produce, wherever they may go. Put differently, with local human capital interventions, place-based objectives and (people-based) returns may be at odds. This said, while we do not negate the issue, we note that much of the development literature often highlights how small internal migration flows are in many developing countries (e.g. Munshi and Rosenzweig, 2016). Factor mobility can also work in the opposite direction. An alternative to the development of local capabilities is the attraction of such capabilities. There is some evidence that attracting facilities from elsewhere through subsidies may generate some effects although the jury is still out regarding the final cost-benefit analysis (Greenstone, Hornbeck, and Moretti, 2010).

[36] An exception is early age human capital interventions which appear to yield higher returns (Cunha et al., 2006 or Almond and Currie, 2011). These interventions are beyond our scope. Typical LED policies target people who have already entered the labor force.

----------Box: 3.5: Human capital externalities ----------

For cities and regions in developed countries, there is a large literature that documents two important facts. First, a higher proportion of university educated workers (or residents) in an area is strongly associated with higher wages, after conditioning out individual characteristics. This last qualification is important. This is a not a composition effect: the skill structure of an area affects the wages of its residents over and beyond their own skills (see Combes and Gobillon, 2015, for a review of this literature). Second, the share of university educated workers in an area is a strong predictor of its future population growth (see Duranton and Puga, 2014, for a review). In both cases the effects appear relatively large. A one percentage point larger share of university educated workers in an area is typically associated with one-half of one percent to one percent higher wages. This is about the same magnitude as the effect of individual university education on one's wage. The same one percentage point larger share in university educated workers is also typically associated with about half a percent higher population a decade later.

Human capital locally may also affect amenities, arguably positively. In this case part of the measured effect on population growth is caused by high-human capital areas being more attractive. At the same time, the productivity effects measured through wages is underestimated as workers are willing to accept lower wages to live in places with better amenities, all else equal. Research by Shapiro (2006) is suggestive that the bulk of the human capital effects measured in U.S. cities are productivity effects rather than amenity effects.

More generally, there is some evidence that these two relationships may be causal. Obviously, it is difficult to find exogenous determinants of the share of university-educated workers in an area to causally identify their effect. For U.S. cities, the literature has relied on the creation of land-grants colleges that were mandated by the 1862 Morrill act to be set up in central locations of many U.S. states and subsequently turned into large public universities (Glaeser and Saiz, 2004, Moretti, 2004, Shapiro, 2006). This also led to the development of college towns that are today highly educated and whose location is plausibly exogenous.

---------- ---------- ----------

Turning to entrepreneurship and enterprise development, there is again research that is strongly suggestive of positive effects of local entrepreneurs on local growth. See box 3.6 for a brief discussion and further references. Again, much of that evidence comes from developed countries and much less is known from developing countries. Support for a positive role of entrepreneurship on subsequent local growth is provided by Ghani et al. (2013, 2014) for India and Duranton and Martin (2018) for Colombia. A key problem for developing countries is data limitations. While India and Colombia track (informal) microenterprises, these two countries are the exception, not the rule. There are also doubts that what has been learned in more advanced countries can be directly used in developing countries. As is well-known in the entrepreneurship literature, everywhere Schumpeterian innovative entrepreneurs coexist with defensive and necessity entrepreneurs (Schoar, 2010, Quatraro and Vivarelli, 2015). However, their proportions may differ considerably in developing countries with large informal sectors. The problem is even worse for innovation given the perhaps very different nature of innovative activity in developing countries.

Just as with human capital, a case can be made that more entrepreneurship may help lagging regions. The issue is then whether entrepreneurship and firm development can actually be fostered in developing countries, especially in their lagging regions. Evidence on the subject is sparse, although McKenzie and Woodruff's (2014) careful review of private sector support in developing countries is not particularly encouraging. The effects of the enterprise development schemes that have been assessed are often insignificant because they are based on samples that are too small. While there are cases of interventions that appear to have made a difference for the treated firms like in the case of the intense involvement of management consultants with a small number of Indian textile firms reported by Bloom et al. (2013), it is difficult to be optimistic in the vast majority of the cases reviewed by McKenzie and Woodruff (2014). After some initial changes in their operations, treated firms often revert to their old ways. Even in cases where firms operate substantially and permanently better after some training, it is unclear whether the economic gains for the treated firms are worth the cost of the intervention as it is hard to track operating profit of firms. Just as with other place-based policies, displacement is also a serious concern. A large growth in revenue of the treated firms may be associated with only minimal productivity gains allowing those firms to expand at the expense of others in highly competitive markets. A training program may even generate negative benefits if the wrong firms are treated.

Beyond these considerations, it is also often unclear which market failure is being solved when providing firms with some free training that they would otherwise not purchase. Self-employed workers may of course be credit-constrained. But then the better response should be to tackle credit constraints in the first place. While we appreciate that developing economies operate deep into the second best, it is nonetheless fundamental to assess the main market failure at play. Besides credit constraints, another possible reason for firms not purchasing training may be a lack of information about the benefits from training. If this is the case, informing firms about the benefits from training may be immensely cheaper than providing this training for free and it may lead to the provision of more relevant training subject to a market test.

From this we conclude that while changes in human capital, entrepreneurship, and firm management in an area can have sizeable effects on final outcomes of interest such as productivity and employment, the literature is not encouraging when it comes to the ability of policies to foster these changes. In a way, our conclusions regarding human capital and enterprise development are opposite to the conclusions we drew regarding corridors. Corridors often generate large quantity responses, but their social value is unclear. Human capital and enterprise interventions may be of a high social value but changing quantities appears extremely hard.

----------Box: 3.6: Entrepreneurship and local growth ----------

We note the existence of a large literature in developed countries that links various measures of entrepreneurship to a number of outcomes such as subsequent employment or productivity growth. Following Glaeser et al. (1992), the positive correlation between inverse initial firm size in a given sector of economic activity and area (or any reasonable proxy of entrepreneurial activity) and subsequent employment growth has been confirmed repeatedly (see Duranton and Puga, 2004, for a review).

Three important caveats are in order. First, small firm sizes may only be a poor proxy for entrepreneurship. Second, employment is not productivity. While job creation may be of social value, positive employment growth in a location and sector may be the consequence of a (relative) productivity decline. This is the case in particular when a good is local and the price elasticity of its demand is less than one. As productivity increases, demand increases less than proportionately. Interestingly, Cingano and Schivardi (2004) find much weakened effects when estimating the correlation between local productivity growth in a sector and initial firm size. Third, using a predetermined variable like firm size to explain subsequent productivity growth in a given area and sector does not make this variable exogenous. Entrepreneurs may enter in more promising sectors and areas. Glaeser et al. (2015) instrument firm size in U.S. cities using proximity to old mines under the argument that mining leads to a more managerial rather than entrepreneurial local culture and find supportive evidence of a causal effect of small firms on subsequent local growth.

---

v) Conclusions

Overall, neither broad-based, big-push type of interventions nor specific human capital or enterprise development instruments appear to be able to make a sizeable difference to the existing, highly uneven equilibria. It is only when massive resources are invested like in the Tennessee Valley Authority program that some significant changes may be observed (though this is far from guaranteed). The TVA involved transfers of up to 10% of local income annually for many years. Returning to the example of the Plan Belgrano in Argentina mentioned above, an intervention of the scale of the TVA would imply annual transfers of about 1.5% of the country's GDP given the economic size of northern Argentina. This is six times as much as the existing plan, which is scheduled to last for 10 years instead of more than 20 for the TVA. The political feasibility of something of the same scale as the TVA for Argentina (and many other developing countries) is doubtful. The wisdom of such an intervention would also be questionable given that the benefit to cost ratio of the TVA is not unambiguous.[37]

We also keep in mind the importance of factor mobility. As already mentioned, raising capabilities may be of social value but the realization of that surplus may happen outside the area of intervention. For workers, higher skills may increase the returns to migration. There is ample evidence that long-run convergence within the United States since the 19th century has been

[37] There are also issues of scalability. The north of Argentina contains nearly a quarter of the country's population. This is an order of magnitude more than TVA counties relative to the United States in 1940.

driven by factor mobility, especially labor mobility, working in conjunction with structural change (Caselli and Coleman, 2001, Michaels et al. 2012, Hornbeck and Naidu, 2014). Put slightly differently, a significant policy effort like the TVA can perhaps be successful at fostering industrialization in a region but the bulk of the convergence of the U.S. South arose from low-skills workers either (i) remaining unskilled but moving away from agriculture, often in another part of the country; (ii) becoming more skilled and moving to cities in the same region or elsewhere; or (iii) staying on the land but enjoying higher returns caused by the changing factor proportions induced by (i) and (ii).

Even in the absence of structural change, it is hard to overstate the role of factor mobility in the adjustments of local economies after negative shocks (Blanchard and Katz, 1992, Hornbeck, 2012). It is also the case that as frictions on mobility arise, convergence stops. Ganong and Shoag (2017) document a link between the drop in regional convergence in the United States and regulations that make housing unaffordable in prosperous places. Unsurprisingly perhaps, interests that benefit from the status quo, however inefficient, resist this type of adjustment just like landowners in the U.S. South resisted black out-migration (Hornbeck and Naidu, 2014).

4. Urban infrastructure and housing

Urban areas are established areas of economic activity and policy is more likely to be catch-up, addressing issues of congestion and inadequate housing, as well as urban expansion. At the same time, the spatial context means that policy is directed at enabling better connectivity between economic agents, and addresses externalities and the consequences of coordination failure. We start with an outline of the urban economy, and then look at policies towards urban infrastructure (section 4.2) and housing (section 4.3).

4.1. The urban economy

As we observed above policies have direct effects, and indirect effects principally via induced private sector investment. Most urban policies are focused on direct effects, such as the benefits of slum improvement or better traffic flows. However, indirect effects also matter. The success of cities in attracting investment and creating jobs has much to do with the way the city has been built – its infrastructure and connectivity, and its housing stock and consequent amenity value to urban workers. The arguments put forward in section 2 of this paper apply with particular force in cities. The costs associated with working and living are high – e.g. hours taken each day in commuting – so the productivity benefits that cities offer must be correspondingly larger, since firms and workers continue to flock to cities. This productivity advantage has been widely researched, and there is a reasonable consensus that the agglomeration elasticity with respect to city population is between 2% and 5%. That is, urban wages increase by 0.2% to 0.5% when city population increases by 10%.[38] Most of this research is done in high-income countries, although research on developing economies indicates that benefits may be even larger (Glaeser and Xiong, 2017). The evidence about urban cost is less developed.[39]

Understanding urban PBP requires that the city be seen as a whole, and hence that the interaction between urban benefits and costs is understood. In section 2 we made the point that the outcome is not, in general, efficient. Private decision taking means that, for the marginal entrant, private benefits are equal to private costs. But external effects created are not taken into account in this decision, so further social net benefits arise according to whether or not the value of positive external benefits created is greater or less than the cost of negative ones. In an urban context, this point is exposited more fully in Box 4.1.

[38] See Combes and Gobillon (2015) for a review of this literature. The apparent elasticity of wages with respect to city population is typically around 8 to 10% but much of that relationship is driven by the sorting of more skilled individuals into larger cities. City population is also potentially endogenous as we expect workers to move to more productive cites. Correcting for these biases leads to smaller elasticities of 2 to 5%. It is also the case that workers may learn more over time in larger cities.

[39] Combes et al. (2017) develop an approach to urban costs grounded in consumer theory and ask what the increase in expenditure needs to be as a city grows in population to keep the same utility. They find that the elasticity of urban costs with respect to population is small but increases sharply for cities with more than a million residents.

----------Box 4.1: Urban equilibrium (from Combes et al., 2005) ----------

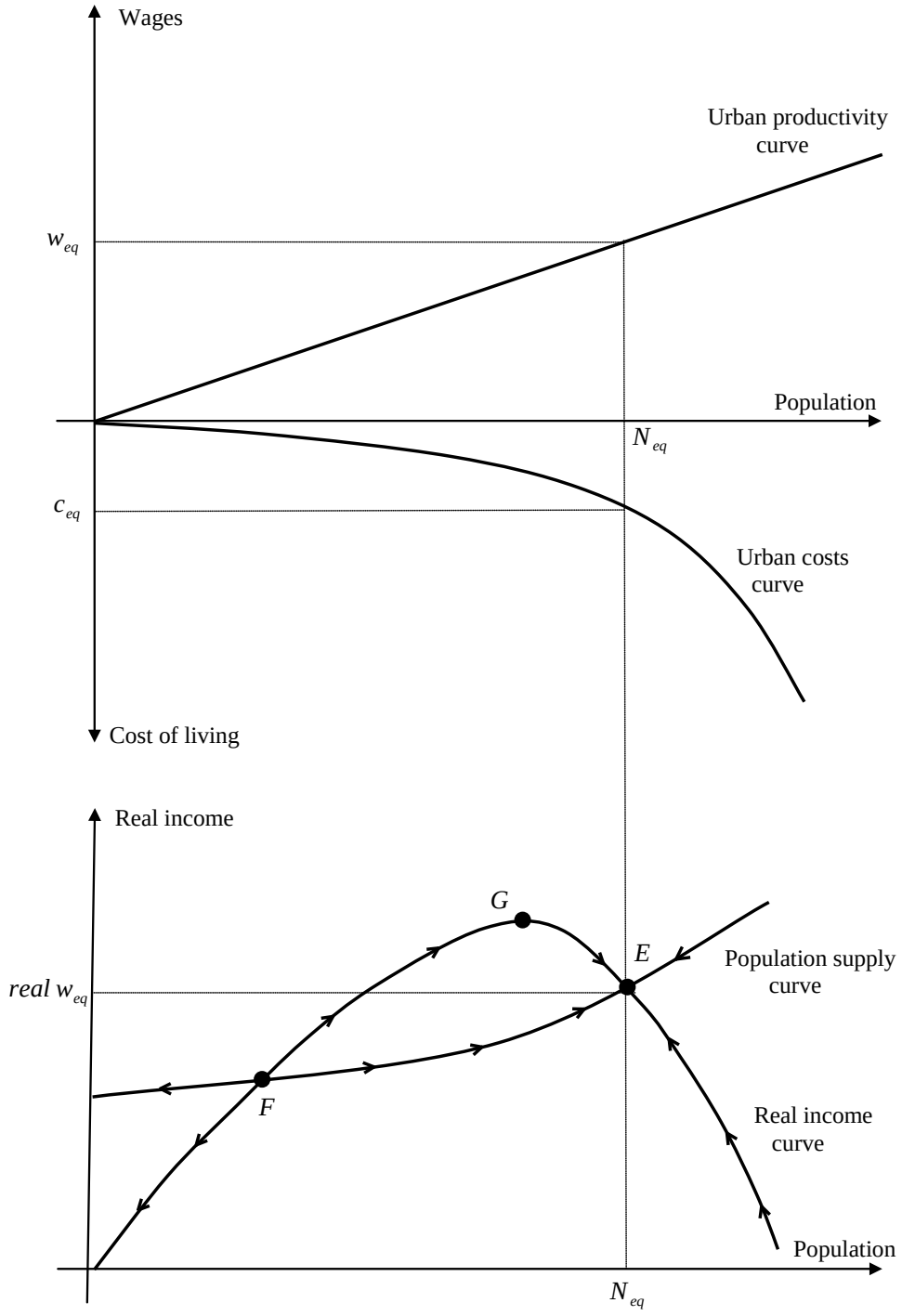
The top section of the figure plots the benefits from agglomeration. The wage of residents increases with city population. This is represented by the urban productivity curve. Think of both axes as being on a log scale. The linearity of urban productivity curve indicates constant agglomeration elasticity, in line with empirical findings (Combes and Gobillon, 2015). The middle quadrant has a reversed vertical axis and represents the rising costs of living in a city faced by residents as population grows. It is convenient to think about this as representing a price of living index in a city. Unlike the urban productivity curve, the urban cost curve is log convex, which reflects sharply increasing costs as cities get large (Combes et al., 2017).

The first curve on the bottom quadrant represents real income in the city, which divides the wage by the cost of living. With logarithmic vertical axes, this curve is obtained graphically by summing the urban productivity curve and the urban cost curve on its reversed axis. The other curve represents the supply of population for the city at hand. This is the real income that residents ask for to reside and work in the city. This curve is upward sloping. Its slope reflects mobility costs and/or a less than fully elastic demand for the city caused by locational preferences. A flat supply curve corresponds to the limit case of perfect mobility. Note that this supply curve implicitly reflects the amenities that the city at hand offers. A city with better amenities would have a lower supply curve given that residents would be willing to forego some of their consumption (real wage) to enjoy better amenities.

The urban equilibrium is point E where the real income curve meets the population supply curve. The corresponding population N_{eq} is such that the real wage that the city can offer (through the real income curve) is equal to the real wage that the marginal resident requires to be in the city (through the population supply curve). We can then read the equilibrium cost of living and wage by reading back on the first two quadrants.

We note that point E is not the unique equilibrium. The population supply curve and the real income curve also intersect at point F, but this is unstable. A small positive shock at F would create a positive wedge between the real income and what the next resident requires to come to the city. This resident would move in, increasing the wedge further. In turn, this would lead to more residents moving in until the stable equilibrium in E is reached. By the same reasoning, a small negative shock in F would lead to the disappearance of the city.

Importantly, equilibrium population is too large relative to point G. This point would allow residents to enjoy a higher real income. This is due to a coordination failure. Should the city be at point G, it would offer potential residents more consumption than they would require to move in as the real income curve at point G is strictly above the population supply curve. Hence, point G cannot be in equilibrium and new residents move in until the two curves intersect. At this equilibrium, the city is too large. However, no single resident has an incentive to leave the city and form a new one since the real income of a new city with a very small population is extremely low. Note that we did not call point G optimal. It is only a constrained optimum that embeds a series of inefficiencies. The next box explains this issue further and derives a full urban optimum.


Wages
Urban productivity curve
w_{eq}
Population
N_{eq}
c_{eq}
Urban costs curve
Cost of living
Real income
G
E
Population supply curve
real w_{eq}
F
Real income curve
Population
N_{eq}

----------Box 4.2: Cities, too small or too big? ----------

As discussed in the previous box, cities have a tendency to be too large because of a coordination failure in city creation. No one has an incentive to move away from an existing city to create a very small city with little to no agglomeration effects. Other inefficiencies are also at play.

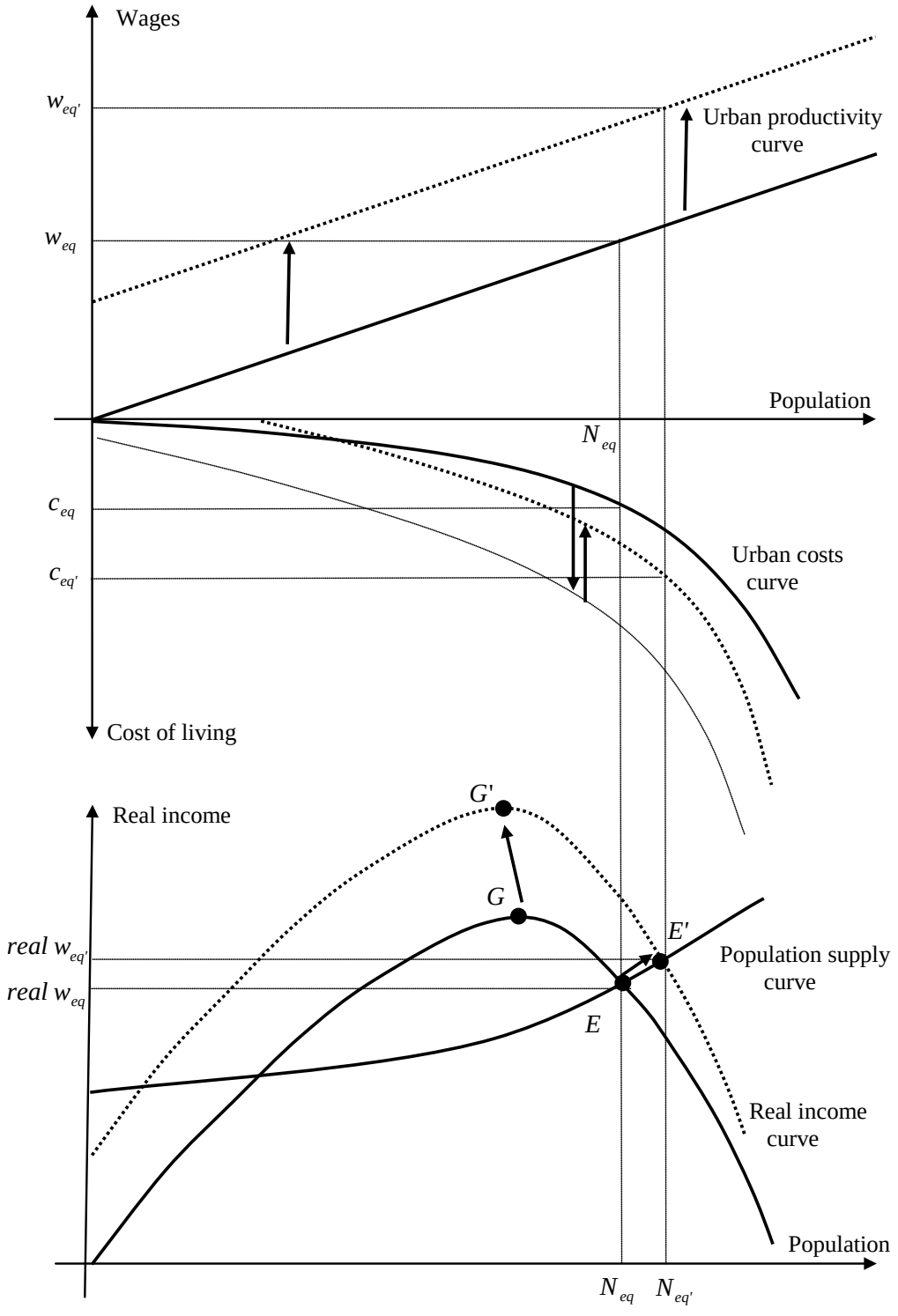
First, when moving to a city a worker considers the prevailing wage in this city. This wage reflects the average productivity of labor, not its marginal productivity since workers do not internalize their own positive agglomeration effects on others. Correcting for this externality shifts the productivity curve up. In turn, this change will also shift the real income curve up. On our graph, this would not affect the level of population that maximizes real income because the urban productivity curve is linear and shifts upwards in a parallel fashion with our log representation.

Second, a similar argument applies to urban costs. In their location decision, workers consider the average cost and not the marginal as they ignore for instance the effect on their choice of residence on traffic congestion. Correcting for this inefficiency would shift the curve and raise the urban costs workers face when moving into a city. We also need to worry about the ownership of land. Having more workers in a city makes land more expensive. Workers perceive this as a cost. Economically, this is only a transfer. Efficiency requires rebating land rents to the local population. This should lower urban costs. These two changes – making residents pay for the marginal rather than the average and rebating land rents – are represented in the middle quadrant of our graph. We also expect local landowners to distort the local supply of land by imposing stricter land use regulations locally. The evidence that, in many U.S. cities, land use regulations are overly strict appears overwhelming (Gyourko and Molloy, 2015). Relaxing those overly restrictive regulations would lead to a flatter urban cost curve. Overall, the efficient urban cost curve may be steeper or flatter than the uncorrected original cost curve. The efficient urban cost curve may also be above or below the uncorrected urban cost curve everywhere or only in some regions.

Turning to real incomes, the efficient real income curve will embed all the changes needed to make the urban productivity curve and the urban cost curve efficient. It will also need to account for the fact that correcting for agglomeration and urban costs externalities requires some transfers that eventually need to be financed locally.[40] In the bottom section of our graph, correcting for externalities leads to a higher real income curve whose maximum is slightly to the left of that of the original curve. It is easy to think of situations where the optimum would be to the right not only of the initial maximum but also of the original equilibrium.

Importantly, making the city "efficient" with regards to its agglomeration and urban costs externalities is not enough to obtain an efficient city population in equilibrium. With our graph, the new equilibrium is at point E'. Real income is higher than previously in absence of correction for the urban externalities but not optimal. Reaching full efficiency still requires solving the coordination failure for new city creation.

[40] Under some conditions the net tax rebate will be zero as it will just offset the net transfers needed to correct for the urban externalities. This result is referred to as the Henry George Theorem. See Arnott (2004).


Wages
$w_{eq'}$
w_{eq}
Urban productivity curve
Population
N_{eq}
c_{eq}
$c_{eq'}$
Urban costs curve
Cost of living
Real income
G'
G
E'
real $w_{eq'}$
real w_{eq}
Population supply curve
E
Real income curve
Population
N_{eq}
$N_{eq'}$

4.2 Urban transport

i) Urban transport improvements: Setting the scene

Before describing the interplay between the demand and supply of urban travel, a number of preliminary points must be made because urban travel is an unusual commodity. Urban travel is not directly consumed. We do not typically conduct errands to get a direct satisfaction from our trips. Instead, urban travel is a costly activity that we undertake in order to produce, shop, or enjoy leisure. The demand for travel is often described as derived demand by transport economists, that is as a cost we pay to enable us to do "things", including work.

We can draw two important implications from this simple observation. First, for a given trip, the total cost of reaching our destination (and coming back) is the relevant economic cost to consider. Then, the total cost of a trip includes both monetary costs such as the gasoline consumed by the vehicle during a trip and the cost of time. In turn, this cost of time will include various elements such as the duration of trip, the discomfort of a particular mode of transportation, traffic conditions, not leaving at our preferred time, and reliability (Small, 2012). Second, we elect to take only a small number of trips from a much broader choice set. We should thus think not only about the trips that we take but also about the entire choice set of the trips that we could take. We can call "accessibility" the combination of the possible destinations and the total cost of reaching these destinations. In short, accessibility measures our ability to go places.

Then, it is convenient to think about the total cost of a trip as the product of its length or distance by its cost per unit of distance. We can refer to the former as proximity and the latter as mobility (or speed if we only consider time). While most of the focus of urban transport policy is on mobility (Duranton and Guerra, 2016), proximity is possibly more important. When they decompose the variance of the duration of a typical trip across Indian cities, Akbar et al. (2017) find that proximity matters more than mobility. This is because trips to specific destinations differ more in terms of average distance than average speed across Indian cities. In addition, accessibility and mobility need not be positively correlated. For instance, denser locations may offer slower mobility but better proximity. The latter effect could potentially be stronger so that accessibility increases with urban density despite worse mobility.[41]

Finally, it is also worth keeping in mind that we pay for travel directly through the cost of travel. We also pay for travel indirectly through higher housing costs in residential locations with better accessibility. This interaction between housing and transport is at the heart of our understanding

[41] For trips to restaurants in U.S. cities, Couture (2015) finds that the time to reach a restaurant from home is roughly invariant across cities. Restaurants are closer in slower cities so that better proximity offsets slower mobility. Denser/slower cities also have many more restaurants and residents choose restaurants that better suit their needs (by passing many more restaurants than in less dense cities). Thus, denser and slower cities have better accessibility.

of land use in cities.[42] We expect land prices to capitalize, at least partly, differences in accessibility across locations within a city.

ii) Urban transport improvements: Price and quantity changes

In a demand and supply diagram, the simplest way to think of an urban transport improvement is to view it as a downward shift of the transport supply curve, that is, a lower cost of travel per unit of distance, all else equal. This situation is represented in box 4.3. The first important implication from thinking about a transport improvement in this light is that the equilibrium cost of travel declines, and this leads to an increase in traveler surplus. The second key implication is that the decline in the equilibrium cost of travel is less than the downward shift in the supply curve. As travel becomes cheaper, it induces more travel, a widely recognized phenomenon usually referred to as induced demand.[43] As travel demand is more elastic, the surplus generated by a transport improvement falls and reaches zero for a perfectly elastic demand which crowds out any change in supply. We return to the interpretation of this seemingly paradoxical result and provide some important qualifications below after our analysis of congestion.

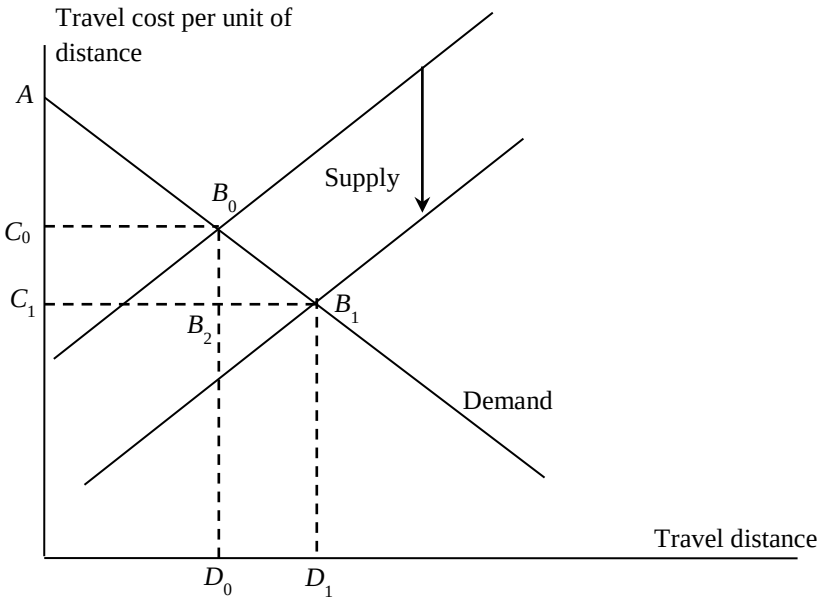
Before going further, it is worth pausing and asking how big the differences in the cost of urban mobility are. The international evidence reviewed by Duranton and Guerra (2016) suggests that cross-country differences in urban travel speed can be large. Travel speed in Mexico City or in Bogota is only about half of that observed in large American metropolitan areas. Similar differences are observed by Akbar et al. (2017) among cities of India. As a result, for a roughly similar share of expenditure and a comparable amount of time devoted to transportation, we find that residents of slow cities travel only half as much or less as residents of fast cities. The economic cost of low mobility appears to be high. Given that households typically devote between 10 and 20% of their income to transport expenses and one to two hours a day per person to travel, the potential gains from better urban transport may be large.

While the framework discussed so far is a useful departure point to think about urban transport, it neglects several important dimensions. The first is that urban travel generates a number of externalities. On the negative side, congestion is the main, but by no means the only external effect associated with transport. Urban travel is also responsible for many road accidents and generates various forms of pollution, from particulates to carbon emissions, to noise. These externalities are discussed by Parry et al. (2007) who make a strong case that accidents and pollution may each be of the same magnitude as congestion in terms of social costs.

[42] See Duranton and Puga (2015) for an extensive review of this literature.

[43] See Litman (2017) for a review.

----------Box 4.3: A simple representation of the supply and demand for travel----------



The horizontal axis measures total distance traveled in a city and the vertical the cost of travel per unit of distance. The demand curve represents the willingness to pay for urban travel, which we reasonably assume to be decreasing in total distance traveled. The supply curve represents the cost of travel as a function of the total quantity of travel undertaken. It is upward sloping because of congestion. Initially, the equilibrium is at B_0 where demand and supply intersect. After a transport improvement such as the widening of a road, the supply curve shifts downwards, and the new equilibrium is at B_1.

Importantly, this improvement leads to an increase in consumer surplus. Originally, traveler (consumer) surplus is measured by the triangle AB_0C_0. After the improvement, it is given by AB_1C_1. This increase has two components. The first is the rectangle $B_0C_0C_1B_2$, which corresponds to the reduction in travel cost for the original trips. The second is the triangle $B_0B_1B_2$, which corresponds to the surplus associated with the increased travel induced by lower travel costs. In practice, these are either new trips or longer versions of previous trips such as going to a more distant supermarket instead of the local corner store.

Transport improvements can also support the positive externalities associated with better access or 'connectivity'. Agglomeration economies are reaped when workers are closer together – employment densities in excess of 150,000 workers per km^2 in the center of major cities – and this density requires a highly effective transport infrastructure. Transport investments are therefore necessary for delivering the high productivity of economic agglomeration.

The next two sub-sections deal with the negative and positive external effects of urban transport in turn. Before turning to these we note two further points. A limitation of the simple demand

and supply framework used so far is that there are several transport technologies, from walking to private cars.[44] Transport technologies affect each other in a variety of ways. More cars and heavy vehicles may discourage biking and other lighter modes of transport. Transit investments such as bus rapid transit with rights of ways will reduce the amount of roadway for other forms of transport, etc. In addition, transit often operates with increasing returns. Hence, easing travel with private vehicles can have detrimental effects on transit, etc. More generally, a cost-benefit analysis of a transport improvement must consider these interactions across transport modes.

Finally, transit improvements in a city are usually far from uniform as they often consist of small projects such as a new bus line or a new overpass or underpass at a particularly busy intersection. In a network, an improvement somewhere can generate a worsening elsewhere. For instance, a new major arterial may impede traffic on connected roads and lead overall to a worsening of total travel time.[45] This type of consideration implies that a good cost benefit analysis should model each transport improvement at a great level of detail. Unfortunately, this imposes extremely high information and modeling requirements for each project.[46]

iii) Urban transport: Congestion

Road congestion is a concern in all large cities, particularly large developing cities. The two simplest ways to think about congestion are the following. First, congestion can be viewed as a friction which worsens with the number vehicles on the road. For security reasons, drivers want to leave a time interval of 1.5 to 2 seconds with the vehicle in front of them. As the density of vehicles on the road increases, the only way to keep this time interval is by slowing down.[47] Second, there are bottlenecks. Intersections and some particularly hard to avoid road segments, such as bridges, have a finite capacity. If vehicles at a bottleneck arrive at a rate that is higher than the rate at which they can exit, a queue will form.

While urban road congestion is extremely salient and receives a great deal of academic and popular attention, one needs to be careful not to exaggerate its importance in the developing world. Akbar and Duranton (2017) find that for a representative set of trips taken by residents of Bogota in Colombia, travel speed only varies slightly less than one to two between the fastest and slowest hours of the day. Looking at traffic conditions in urban India, Akbar et al. (2017)

[44] While an overwhelming majority of trips in the United States rely on privately-owned vehicles, rates of car ownership are still low in emerging cities. While transit is widely available in large European cities, it is still sparse and often informal in much of the developing world.

[45] This extreme situation is referred to as Braess' paradox.

[46] Transport models often provide a disaggregate modeling of an entire city divided into many small zones. This requires knowing about the demand for trips for the entire matrix of origins and destinations and the modeling of congestion effects on the supply side. Because much of that information is usually missing, it is generated from knowledge of the location of residents and jobs and from relationships such as a gravity pattern for commutes.

[47] Related to this, minor traffic shocks such as small decrease in speed to make way for a vehicle that changes lane gets amplified in dense traffic as each vehicle needs to slow down as least as much as the vehicle in front of them.

find that travel delays caused by traffic only play a minor role in explaining mobility differences between Indian cities. It is only when attention is restricted to the central parts of the very largest Indian cities that traffic delays become an important component of mobility. Among Indian cities, the first-order issue is slow mobility in absence of congestion. Congestion only affects a small number of locations, albeit of disproportionate economic importance.[48]

As shown in box 4.4, congestion can easily be incorporated into the framework used so far. The usual recommendation from economists can often be summarized in one word: pricing. In practice, the easiest form of pricing is to implement a cordon around the central part of a city with a toll to be paid when crossing this cordon at certain hours of the day. Politically, this is difficult to implement since congestion pricing leads to a number of distributional conflicts, between rich residents with a high value of time and poor residents, between suburban residents who are more reliant on cars and more central residents who are often exempt from the charge altogether, and between renters and owners as a congestion tax is likely to affect property prices. We also keep in mind that taxing congestion raises a number of implementation and enforcement issues that may be challenging for many cities in the developing world.

While direct congestion pricing may be out of reach for many developing cities, other pricing mechanisms are possible. Pricing for parking is one of them. Appropriately implemented, the pricing of parking has the added benefit of freeing up parking spots and eliminating cruising, arguably a major source of congestion in many central cities (Shoup, 2005).[49]

Beyond pricing, it is also important to keep in mind that traffic conditions are also greatly affected by driver behavior and the quality of the roadway, ranging from the absence of potholes to well-functioning traffic lights or, more advanced but still easy-to-implement devices, such as ramp-metering systems. In many parts of the developing world, the division of the roadway between moving vehicles, parked vehicles, and other users such as street sellers is not well defined or enforced. Akbar et al. (2017) find that the evolution of travel speed during the day is highly consistent with road encroachments being a major impediment to mobility in Indian cities.

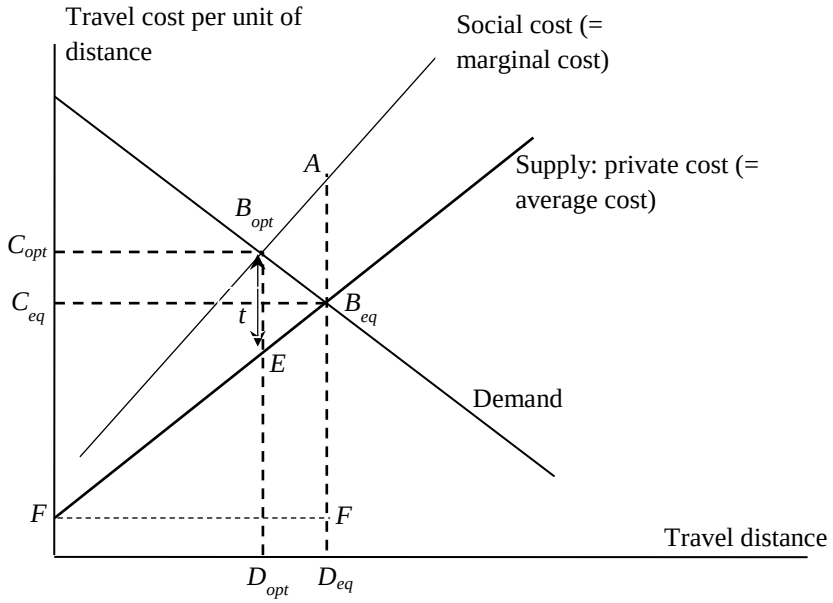
The endless debates about urban congestion also tend to forget that the optimal congestion tax is conditional on the extent of the roadway. Cities with more roadway in India also appear to enjoy better mobility (Akbar et al., 2017). Many emerging cities where informal residential development occupies a large fraction of the land appear to have fewer roadways than equivalent cities in developed economies.

[48] Although we focus mostly on road congestion, it is not the only form of travel congestion. Rail and bus may also get highly congested with congestion costs that are paid in discomfort and long waits in stations.

[49] Rather than price instruments, some cities use quantity restrictions, typically based on plate numbers. These are particularly popular in large cities in Latin America. The jury is still divided on the effects of these measures. Carillo et al. (2016) find that it led to some reductions in travel in Quito. Davis (2008) finds no effect for Mexico City.

----------Box 4.4: Congestion: A graphical analysis----------

The demand and supply curves are as in box 4.3 and the equilibrium is at B_{eq}. The supply curve reflects the private cost of travel per unit of distance, upward sloping because of congestion. However, this does not internalize the externality created by additional travel, so the full social cost is the (higher) marginal cost of travel. There is thus excess travel at B_{eq} as travelers do not pay for the full cost that their travel inflicts on society. The optimum is instead at point B_{opt}, where the marginal cost of travel intersects with the demand curve.



The optimal amount of travel is D_{opt}. For residents to be willing to travel the optimal quantity D_{opt}, they need to face a cost of travel C_{opt}. This can be achieved by taxing travelers and charging them the difference between the marginal cost of travel and the average cost, t.

Without a congestion tax, excess travel at D_{eq} generates a social loss (deadweight loss) equal to the difference between the social cost of travel and the willingness to pay of travelers between the optimum and the equilibrium. In the graph, this loss is measured by the triangle $AB_{eq}B_{opt}$.

The "cost of congestion" is often measured as the total delay relative to a free flow situation. In our graph, that would be the rectangle $C_{eq}B_{eq}FF$. Although seemingly intuitive, this measure is problematic in three respects. First, getting the equilibrium amount of traffic to flow at the speed that would be achieved in absence of traffic is a physically impossible counterfactual. Second, even if free-flow speed could be achieved, demand would not remain at D_{eq}. It would increase. Third, optimal mobility is not free-flow mobility as implicitly assumed when measuring total delay. Optimal mobility involves instead some congestion and the objective is to reach an optimal amount of congestion, not to eliminate it completely. Eliminating congestion completely also implies eliminating travel altogether.

----------Box 4.5: Three further issues about congestion----------

While the graphical device used is box 4.4 is useful and informative, it neglects three important points about congestion.

First, there is good empirical evidence that for a given road, the supply curve is first upward sloping as represented in the figure of box 4.4 but then eventually bends backwards. The reason is that as the number of vehicles increases, their speed keeps decreasing and eventually the overall flow or quantity of travel declines. In the extreme, too many vehicles on a road bring it to a standstill. This phenomenon is known as hyper-congestion. The social costs associated with hyper-congestion can be extremely large. For instance, a hyper-congested highway that moves at 15 kilometers per hour could perhaps achieve the same flow at a speed of 60 kilometers per hour. In this case, 75% of the cost of travel is (deadweight) lost over and above the standard losses from congestion as illustrated in box 4.4.

Second, the literature typically finds that specific roads can be highly congestible with an extremely steep supply curve that can bend backwards (Small and Verhoef, 2007). On the other hand, studies that measure congestion at the area level find that the supply curve is much flatter (Akbar and Duranton, 2017). The reason is that, as major arteries get congested some travelers start using alternative routes. As a result, a given area will be less congestible than a single major road. If there is enough supply of local roads, the supply curve will eventually reach a plateau and become flat as all the new traffic is directed towards uncongested local roads. The optimal congestion tax for the area is then zero. This situation may nonetheless be deeply inefficient as the major artery may be hyper-congested. This situation would then call for the taxation of specific roads instead of the usual “cordon” pricing of an entire area as encountered in London, Singapore, or Stockholm.

Third, we often think of congestion in static terms and the main loss is time lost in traffic. Taking a more dynamic view, we know that the demand for traffic varies a lot throughout the day. A key margin of adjustment to costly travel is re-scheduling. In dynamic models of congestion, travelers face a key trade-off between arriving on time but face a high cost of travel or arriving at a less preferred time but enjoy a lower cost of travel. In the benchmark dynamic model of congestion developed by Arnott et al. (1993), this cost of scheduling delays represents half the total cost of congestion. While this particular 50-50 split depends on the details of the assumptions made in this model, there is little doubt that scheduling costs represent a large unobserved fraction of congestion costs. While traffic jams are extremely salient, scheduling delays are often forgotten.

Re-routing and scheduling delays are useful to keep in mind when interpreting the phenomenon of induced traffic. We often observe that roadway expansions quickly “fill up” and traffic conditions revert to their initial situation at peak hour. This is often interpreted as if these expansions were of no social value. This neglects the fact that the extra traffic generated substitutes for traffic on slower alternative routes or may allow travelers to travel closer to their preferred time.

---

iv) Urban transport improvements, economic centers and land use change

Although much of the discussion about the externalities of urban transport focuses on the negative, there are also positive externalities associated with better access. First, a reasonable first-order conceptualization of emerging cities is to view them as labor markets for their residents. A transport improvement in a city such as the development of a bus rapid transit system will change how residents can access jobs. In turn, this will likely change how employers and employees match. Employees may gain some bargaining power because of more job opportunities. Firms may be able to find better matches from a larger pool of workers. Transport improvements can also ease other daily errands such as going shopping. In turn, more shopping options for shoppers may lead to greater price competition among retailers. Better shopping accessibility can also lead to a greater concentration of retail, a source of potential efficiency in an industry characterized by increasing returns. More generally, transport improvements can have wider economic benefits that take the form of agglomeration effects (as discussed in section 3.1) arising from better access. These positive externalities associated with transport improvements are arguably more important in developing countries where the inability to travel more than a couple of kilometers by foot is a limiting factor for a large share of urban residents.

Capturing these effects can be thought about in two stages. The first is that, even if land use remains unchanged (so no new investment in jobs or housing changing the shape of the city), then a transport improvement makes places closer together, in economic terms, so increases the effective size or density of a cluster of activity. It is possible to use these estimates to produce an estimate of the productivity impact of a transport improvement. The approach has the advantage of capturing the fact that transport enables more workers to get into a cluster, and consequent possible productivity effects, and is routinely used as part of the UK's transport appraisal process (see Box 4.6).

Holding land-use unchanged is a simplifying assumption, particularly since land-use change and transport improvements are often combined in a joint policy package. While the main short-run effect of transport improvements is on the travel behavior of residents, in the long-run they will also affect the location choices of residents and firms. These changes will of course affect how we should value urban transport improvements. In conjunction with this, changes in property prices are often used as a metric to evaluate urban transport improvements. To remain concrete, consider a new bus-rapid transit (BRT) line that links the center of a city to its remote periphery, such as the new BRT under construction in Beirut, Lebanon. In the short-run, this project will lead to an increase in the surplus of residents through cheaper and more travel (assuming that it will indeed lower the cost of travel per unit of distance in Beirut). This project also aims to lessen the dependence on cars of local residents and rebalance travel between transport modes. This is expected to generate a decline in the cost of congestion for the city.

----------Box 4.6: Wider economic benefits and transport appraisal----------

UK cost-benefit appraisal of transport projects is based on user-benefits (and a generalized version of the rule-of-half) and adds several sources of 'wider economic benefit', one of which is a productivity effect created by increasing the 'effective density' of economic activity. The approach is an attempt to combine direct (user-benefit) and indirect (agglomeration economies) effects in a rigorous framework, subject to rigid guidelines set by the UK Department for Transport.

The method is based on two steps: Transport → effective density → productivity.

The first stage employs a formula $ED_i = \Sigma_j f(d_{ij})Employment_j$ which computes the effective density (*ED*) of each place *i* as the sum of employment in all areas, weighted by the reciprocal of a measure of economic distance, d_{ij}. The second step uses a relationship $P_i = (ED_i)^{\alpha}$ where P is productivity and α is the elasticity of productivity with respect to effective density. This is drawn from econometric studies which seek to capture agglomeration externalities and has value approximately 0.04. A transport improvement reduces d_{ij} and hence (given employment) increases the effective density of effected places. This raises the productivity of *all* workers in the affected area according to elasticity α. The productivity increment is a welfare gain that is added to the user-benefits derived from the transport improvement.

Even without any land use change, residents may re-sort within the city. More specifically, residents that depend more on public transport may choose to relocate close to the new BRT station. This resorting of residents leads to serious complications when assessing the value of the BRT project. Resorting may imply that any increase in transit trips observed with the new BRT may be just a substitution for transit trips that were taken elsewhere before, rather than residents switching from car to transit and relieving congestion. This does not imply that the project is not socially worthwhile. Residents that now have easy access to the new BRT may have previously used other forms of transit that were further from their residence, slower, and less frequent. This is a clear benefit to them, but it is not the same as a social benefit from reduced congestion.

Then, this BRT project may lead to changes in land use with, for instance, higher residential densities close to the new stations near the center and new residential developments towards the end of the line. Easy access to transit for an increased number of residents is of obvious value to these residents. For residents that relocate to new residential developments at the periphery, the calculation is more complicated. Their total transport cost may actually increase. These residents in newly developed peripheral areas may be willing to incur higher transport costs through longer distances because it allows them to consume more housing at a lower cost. While these are private gains for these residents, the calculation of these gains is extremely challenging.

A new transit line may not only affect the residential "origins" of trips. It may also change the destination of trips, workplace or other. For instance, with better transport, some firms may decentralize and relocate to the periphery of the city. Retail may also relocate from the center of

the city towards the periphery. This can potentially increase travel distances to access work or retail. In extreme cases, accessibility can even be reduced when the increase in distance more than offsets the reduction in travel cost per unit of distance. When measuring the value of projects, accessibility and not mobility is the relevant welfare measure as argued above.

Because these changes in location and travel behavior are extremely hard to measure and value, a metric that would capture all these changes at once and could be used as a sufficient statistic to measure welfare changes is alluring. The change in aggregate land values is the obvious candidate here. As already mentioned, a lot of caution needs to be exerted when attempting to measure welfare through land values. A first limitation arises from imperfect mobility. For instance, the social worth of a transport improvement in an area will not be fully reflected in land values if potential newcomers face some frictions to move into the area. To take an extreme example, if a city is closed to new residents, a transport improvement will typically lead to lower land values as the accessibility premium from more central locations declines. Alternately, local residents may restrict access to newcomers through restrictions on land use and new developments. Aggregate land value will then reflect their "monopoly power". As transport improvements affect the behavior of incumbent residents and may lead to them to impose more stringent restrictions, the change in land values will not in general reflect the social worth of the improvement.

The second limitation arises from residential heterogeneity. To take a simple example, the potential resident who values a transport improvement the most only needs to pay a little bit more than the potential resident with the second highest valuation to be able to buy the house that benefits from better access. The increase in land value thus fails to reflect the full social value of the transport improvement. The third limitation is that there might be changes in lot size or housing consumption per capita. The change in the welfare of residents should be measured by both the changes in prices of housing or land and their change in the quantity consumed. Finally, general equilibrium effects must be kept in mind. A BRT project in a secondary city in China is unlikely to change welfare in other Chinese cities. A BRT project in Beirut, the dominant city in Lebanon, will affect other Lebanese cities. Increases in land values in Beirut may be partly offset by declines in other cities. This same "growth vs. displacement" problem plagues the evaluation of regional corridors as discussed above.

4.3. Housing and other urban infrastructure

Housing policy in developing countries often follows similar goals as in more developed economies and seeks to improve housing quality, increase the ability of residents to consume housing, and facilitate housing transactions to own or rent housing. The challenges are of course immensely more difficult in poor countries where housing informality is the default choice of a

large fraction of the population, where housing finance mechanisms are often primitive if not absent, and property rights are often disputed, badly registered, and poorly protected. We also keep in mind that a house is not only a roof but should be considered more broadly together with a number of complementarity services and utilities such as water, sewers, and power as well as access to jobs, goods, and schools.

i) Quantity effects: Fostering supply and demand vs. enabling markets

A key complication with housing policy is the vast array of instruments available to policy makers. More "fundamental" institutional policies seek to enable the housing market to work. More direct policies either provide housing on the supply side or allow residents to consume more housing through subsidies on the demand side. In between, and at an intermediate level, the provision of key infrastructure and utilities or appropriate urban planning will hopefully raise the quality of the housing being supplied.

While a huge housing shortfall cannot disappear overnight, housing stocks should be able to adjust reasonably fast to demographics and income changes, provided property markets are allowed to function. In principle, the residential construction industry should face minimal barriers to entry and be competitive while employment in the sector could potentially grow fast. Given this, the provision of a working institutional framework that allows the expression of supply and demand mechanisms to play their role seems like an attractive proposition.

Institutional reform for the land and housing market is unfortunately harder than it may seem at first sight. Many complementary factors need to be in place at the same time. First, property rights need to be well defined. They are sometimes absent.[50] In some parts of the world, property rights often collide with other rights, such as those given to tenants. For land and housing markets to work, property rights also need to be tradable. This arguably requires a reasonably uniform system of rights. This condition is clearly not met in many countries where land rights are administered by customary chiefs, sometimes in a whimsical manner; many systems of right coexist uneasily; and the trade of property assets is heavily restricted within a group (Durand-Lasserve et al., 2015). Then, land parcels need to be clearly delineated and registered through a system of cadaster and land registry. Such a system can only operate if titles are not contested and the registry is kept up to date. Although this latter condition sounds trivial, keeping land registries up to date is a struggle in most developing countries, including some of the more advanced ones. Finally, even when all those conditions are met, the land market will not work unless titles can be enforced and secured by a fair and timely judiciary system (Glaeser et al., 2016). With only a little exaggeration, one may conclude that well-functioning land and housing markets essentially require a high level of development.

[50] Among many examples, Kibera, the largest slum of Nairobi (see Henderson et al., 2017).

The second key set of institutions needed for functional property and housing markets is a system of housing finance. Housing is not only an important good in household consumption; it is also part of an extremely large asset class. In the absence of a dedicated system of housing finance, very few households can easily pay for an asset that is typically worth several times their annual income. Renting is the obvious alternative tenure choice. It does not require households to make a large investment. At the same time, a property on the rental market still needs to be financed by someone and this someone is also likely to need external finance. In addition, rental arrangements are subject to difficult agency issues with potential abuse on both sides. These agency problems impede the development of a rental market in countries with weak institutions.[51] If anything, observation suggests that rental markets are often embryonic in poor countries. For instance, the homeownership rate is above 80% in India instead of the 60 to 70% in most developed countries or even 50% or less in Germany or Switzerland.

Housing finance relies on several sets of conditions, with again extremely limited substitutions between them. The first is a system of titles and a workable market for land and properties as already discussed. The second is a set of laws and regulations for mortgages. These regulations must protect banks from fraudulent borrowers by allowing them to control titles. Regulations must also protect borrowers against potential abuse by lenders and make sure their titles are returned when the mortgage is paid. With a mortgage, both parties enter a long-term relationship that requires some guarantees and predictability. The third set of conditions regards the creation of funding flows that allow lenders to lend to a large number of borrowers. In turn, this requires either large, well-funded retail banks, a system of specialized banks like building societies associated with savings, or a large provident fund with an independent source of funds, such as a payroll tax.

While these more fundamental policies that allow property and housing markets to function and housing to be funded are eventually necessary at higher levels of development, they seem unlikely to fix housing consumption shortfalls in the short- to medium-run in poorer countries. As a result, these countries often turn to interventions that are more direct. Even with ill-functioning property markets, it is possible to develop a large quantity of new housing in a fairly short period of time, as evidenced by mass housing programs in Mexico, Brazil, or South Africa (Buckley et al., 2015). These programs seek to foster both supply and demand by mandating new large housing developments and by offering deep discounts for the newly constructed housing units (which are sometimes given free). Other, less ambitious, policies may act only on the demand side by offering subsidies for house buyers / borrowers or only on the supply side, for instance, by building public housing directly or by offering fiscal incentives for housing investments.

[51] To take only one simple example, a small-scale owner needs a title to be able to collect rents. Otherwise, the tenant could claim the property.

Some of these policies have been successful at increasing housing quantities (though not at creating social value as argued below). For instance, the "My House, My Life" program in Brazil was able to develop nearly two million new houses benefiting about 7 million Brazilians in about five years. Other countries such as Angola, Ethiopia, or Thailand have been far less successful at massive housing development. Their programs have often struggled to build more than tens of thousands of new housing units. The two key factors behind the ability of some of these programs to produce new houses on a massive scale appear to be (i) political prioritization and (ii) the existence of provident funds to fund these constructions.

ii) Quantity effects: Upgrading vs. retrofitting vs. expansion

Before turning to the evaluation of new housing units, the other key quantity decision regards the type and location of new housing. At one extreme, there is urban expansion occurring through new housing built on "greenfield" suburban developments. At the other, there is rebuilding or upgrading in already developed areas, typically slums. Various forms of infilling construction can occur between these extremes.

We need to draw a strong distinction between rebuilding and upgrading. Retrofitting (or rebuilding) entails very significant changes to an area. Land in slums is typically highly fragmented and mostly used for residential purpose with low-rise and low-quality housing covering most of an area. Retrofitting typically involves the destruction of the existing housing stock and very significant changes in land use with a drastic reduction in the share of land devoted to residential purpose to make room for more roadway, commercial development, public services, and recreational spaces. Upgrading involves instead more marginal interventions centered on current residents, including the provision of basic infrastructure and utilities or subsidies and technical help with housing improvements or expansions. Some policies also involve the legalization of hitherto illegal settlements to give them titles (De Soto, 2000), which could be used for a variety of (entrepreneurial) purposes or, simply, access to utilities and public services (since being legal is often a requirement to receive public services).

Distinct from retrofitting and upgrading, urban expansion is another option. Urban expansion is often frowned upon as it involves new developments beyond the urban fringe. It is often disparagingly labeled as "urban sprawl" and accused of a wide variety of social ills. It is true that new suburban housing developments require complementary utility and infrastructure development in areas that are not currently serviced. This is a cost. This type of development is also arguably subject to severe market failures.

The inefficiencies associated with urban expansion are reminiscent of those associated with urban size described in box 4.2. Coordination failures are rife with urban expansion. Not knowing how and where expansion will take place makes developers reluctant to invest. This potentially leads to too few new developments. In addition, there are also many externalities at

play. For instance, new suburban developments in emerging cities often take place without much consideration of their transport and congestion implications. In addition, these developments are often targeted at rich households willing to live at much lower density and are unlikely to factor in all the social costs of suboptimally low density.

Despite the challenges, urban expansion is a pressing issue. Consider the typical case of cities that grow in population by 2% per year and where households are willing to increase their housing expenditure by 2% per year.[52] To prevent prices from increasing, this 4% annual increase in demand implies a doubling in the amount of developed floor-space of housing every 18 years.

Hence, despite its lack of popularity and the costs highlighted above, urban expansion will nonetheless need to be a big part of accommodating increasing demand for housing. The potential for large increases in housing supply from slum upgrading policies seems limited given extant construction technologies and materials being used.[53] Urban retrofitting may not be much more promising in this respect either as much higher constructions in modernized areas may not do more than offset the much lower share of land devoted to housing (Henderson et al., 2017). As we argue below, slum upgrading and urban retrofitting have some virtues but accommodating massive growth in demand is not one of them.

Besides its limited ability to provide a large increase in housing floor-space, urban retrofitting has other drawbacks. The first is its cost. While candidate neighborhoods for retrofitting offer potentially good accessibility and thus have valuable land, tearing down, redesigning, and rebuilding is arguably much more costly there than for greenfield development. The second main drawback of urban retrofitting is that it entails the relocation (at least temporarily) of large groups of residents. While land in some areas seems so valuable that existing inhabitants could be re-accommodated locally after retrofitting without affecting the economic viability of such project, existing experiences are not always encouraging. For the emblematic case of the Dharavi slum in Mumbai, Iyer et al. (2011) show that a viable re-development project supported by the municipal authorities got derailed as poorly defined property rights led to a much larger number of claimants than expected for the projected new housing. This jeopardized the economics of the

[52] Using national accounts for a large cross-section of countries, Dasgupta et al. (2014) show that housing investment as a fraction of GDP per capita is S-shaped and takes off at about $3,000 before tapering off at around $36,000. Low-income countries invest about only half what upper-middle-income countries invest as a fraction of their GDP. Housing is often acknowledged to be a normal good, but it appears to be a luxury good at low levels of development so that a 2% increase per year for housing expenditure at the household level is perhaps a conservative estimate in a poor but growing economy. Then, a 2% per year population growth is above what we have observed in the recent past in Latin America but seems very conservative for large African cities and some parts of Asia, including India.

[53] The most primitive forms of housing are shacks for which building up is simply not an option. With less extreme poverty, houses made of brick can be raised by several floors without requiring full reconstruction.

project as it was feasible to rehouse existing inhabitants in new developments one square meter for one, but not four or five for one. The other main cause behind the failure to redevelop Dharavi is that the proposed plan was opposed by the powerful local leather and dying industries, which, for obvious public health reasons, were scheduled to be relocated elsewhere after redevelopment. In many slums, there are also politically well-connected "slumlords" that benefit from slums through rents that they collect either directly from the housing that they control or from the services and utilities that they charge to residents.

Slum upgrading and legalization offer a third, seemingly more modest option. Providing titles is sometimes viewed as a magic wand for development, giving previously asset-free residents collateral that could allow them to start new businesses (De Soto, 2000). Recent academic studies of titling experiences have failed to provide evidence on this collateral aspect. They point instead at a more modest reality where titled households increase their investment in housing and in their children's education (Galiani and Schargrodsky, 2010) or their labor market participation (Field, 2010). Given these results, titling policies appear to be a useful tool that belongs to the broader family of slum upgrading policies together with improvement in water and sewers, and financial and technical support for housing expansions. An important downside of slum-upgrading policies is that they appear to slow down the "modernization" of the neighborhood where they take place (Harari and Wong, 2017).

Despite its costs, its adverse political economy, and the potential for social upheaval associated with the relocation of residents, urban retrofitting will eventually occur in most emerging cities. It occurred in all large cities of what is now the developed world. The question should thus not be whether slum redevelopment will take place, but when and how. In turn, this begs the issue of the articulation between urban expansion, retrofitting, and slum upgrading. It looks like slum upgrading is needed but is more of a stop-gap policy. While slum upgrading is possibly of value, its negative effect on retrofitting needs to be factored in. Then, there are some potential complementarities between retrofitting and urban expansion. The choice here is not either / or. Both are arguably needed as existing urban boundaries will have to be pushed outwards to accommodate demographic growth and the rise in demand for housing in most emerging cities. At the same time, good locations that have been informally developed should not remain with highly adverse patterns of land use forever. As already mentioned, retrofitting is extremely challenging, in part because of extremely high human densities in slums. Urban expansion that allows some informal households to relocate in more formal and newly developed areas will put some relief on the demand for slums and, in turn, possibly ease their conversion.

Finally, we note that these policies need to rely on a variety of instruments. Some of these instruments are hard instruments such as the provision of complementary infrastructure and utilities without which new construction would be unlikely to take place. There are also soft instruments, including zoning designations, expansion plans, etc. Both types of instruments are

needed. Developers are unlikely to start building a new residential complex if it cannot be accessed and serviced with utilities. They are also unlikely to build if they do not know what to expect for the immediate surroundings of their parcel. We note that both soft and hard policy instruments also require some time consistency, as housing investments are extremely long-lived and unmovable. More concretely, it is important to plan for future roads. It is also important that this plan is credible, as squatters may otherwise occupy this land and develop it informally. The durability of housing also implies that policies must get their timing right. If we take the example of providing utilities to newly developed areas, it is easy to understand that doing it too early is wasteful and may lead to an infrastructure being depreciated before it is fully used, whereas doing it too late will imply costly retrofitting costs.

iii) Valuing increases in housing consumption

Many housing policies, especially on the demand side (like direct subsidies) come at a high cost. More fundamentally, these policies are also often questioned because housing is a standard private good whose consumption is both rival and excludable. Subsidizing housing when residents would prefer to consume other goods may entail a large deadweight loss.

We can distinguish several types of benefits for housing policies. First, we can think of the more fundamental policies about property rights as reducing frictions to exchange. For instance, when housing is informal it is extremely difficult to trade. This implies that residents may be stuck in housing that is extremely far from their workplace. Improving the functioning of housing and property markets can thus lead to better allocations. Although such benefits may not be salient, they are potentially large. The relative certainty offered by titles and the option to resell a property if needed also appear to foster the incentives of households to invest in their accommodation.

Second, households are generally financially constrained, particularly so in poorer countries. Although the literature on titling has failed to demonstrate that households are willing to use their title as collateral in entrepreneurial ventures, there is overwhelming evidence that they are willing to do so to obtain a loan to purchase a house and increase their housing consumption.

Third, while housing is generally a private good, residents may not factor in health and schooling benefits from less crowding and generally better housing. It is also the case that the provision of housing services does not result only from having a roof over one's head. It also results from complementary investments in utilities and access. Water, sewers, and roads are largely public goods and significant externalities are associated with them. There are potentially large public health externalities associated with the provision of water and sewers (Ashraf et al., 2015), and it appears that urban dwellers in emerging cities value the pavement of their streets highly. (Gonzalez-Navarro and Quintana-Domeque, 2016). The same high valuations probably hold true for many other urban utilities and public goods.

Transport and access seem to play a particularly important role in this respect. Simply put, housing without access is worth very little. The mass housing programs described above have been major failures in this respect. Although countries like Brazil, Mexico or South Africa have been able to expand their number of housing units very significantly, this has come at a large social loss. Housing units developed by these policies typically re-sell at a price well below their cost of constructions and some large developments have remained mostly empty (Buckley et al., 2015). The failures have been multiple, including the insistence of these policies to develop large areas zoned exclusively for residential purpose without retail or any form of economic activity. The largest failure has been perhaps the disregard of any notion of accessibility and especially of accessibility to jobs. With a fixed budget per unit, developers built new housing in areas where land could be bought essentially for free to save on costs. There are of course good reasons when land is essentially free, and lack of access is usually prominent among them. Housing is place-based. This is unfortunately too often forgotten by housing policies.

5. Concluding comments

The outcomes of place-based policies (PBP) are inherently uncertain, but choices have to be made. This paper seeks to inform these choices in two ways.

First, it provides a framework in which analysts can identify, organize, and assess the information needed to make choices. The approach is grounded in economic principles and emphasizes two elements. One is understanding the mechanisms through which policy may bring about change (quantity effects), and the importance of looking beyond immediate (or direct) impacts of a policy to induced changes in private sector behavior throughout the economy. The other is the importance of valuing changes accurately, essentially by clear thinking about opportunity costs, the value of resources in alternative uses, and the market failures and inefficiencies that motivate the policy intervention.

Second, the paper lays out some of the issues that are particular to the spatial dimension of place-based policies, in general and for particular policies. This is a context in which there is no simple mapping from policy action to outcome. Increasing returns to scale, coordination failure, and the fact that multiple conditions have to be satisfied for policy to be effective are all sources of difficulty. The paper reviews the issues and provides pointers to the research literature on what is known about the effects of PBP. This is a research literature that faces its own challenges, both in identifying the effects of policy change and in the ability to generalize across to new areas and projects.

This paper and the literature it draws on do not offer a comprehensive body of experience of comprehensive rules for appraisal of PBP. Nevertheless, a simple but rigorous framework will, we hope, provide a basis for more informed discussion about projects and for better decision taking and outcomes.

References:

Aghion, Philippe, and Peter W. Howitt. 2008. *The Economics of Growth*. Cambridge (MA): MIT press.

Almond, Douglas, and Janet Currie. 2011. Human capital development before age five. In Orley Ashenfelter and David Card (eds): *Handbook of Labor Economics, Volume 4*: 1315-1486, Amsterdam: Elsevier.

Akbar, Prottoy A., Victor Couture, Gilles Duranton, Ejaz Ghani, and Adam Storeygard. 2017. Accessibility and Mobility in Urban India. World Bank Working Paper, forthcoming.

Akbar, Prottoy A. and Gilles Duranton. 2017. Measuring the Cost of Congestion in a Highly Congested City: Bogota. Mimeographed, University of Pennsylvania.

Angel, Shlomo, Jason Parent, Daniel L. Civco, and Alejandro M. Blei. 2012. *Atlas of Urban Expansion*. Cambridge, MA: Lincoln Institute of Land Policy.

Arnott, Richard. 2004. Does the Henry George Theorem Provide a Practical Guide to Optimal City Size? *American Journal of Economics and Sociology* 63(5), 1057-1090.

Arnott, Richard, André De Palma, and Robin Lindsey. 1993. A Structural Model of Peak-Period Congestion: A Traffic Bottleneck with Elastic Demand. *American Economic Review* 83(1): 161-179.

Ashraf, Nava, Edward Glaeser, Abraham Holland, and Bryce Millet Steinberg. 2015. Water, Health, and Wealth. Work in progress, Harvard University.

Azariadis, Costas, and Allan Drazen. 1990. Threshold externalities in economic development. *Quarterly Journal of Economics* 105(2): 501-526.

Baldwin, Richard, Rikard Forslid, Philippe Martin, and Frédéric Robert-Nicoud. 2005. Economic geography and public policy. Princeton (NJ): Princeton University Press.

Berg, Claudia, Uwe Deichmann, Yishen Lieu, and Harris Selod. 2017. Transport Policies and Development. *Journal of Development Studies* 53(4): 465-480.

Blanchard, Olivier J. and Lawrence F. Katz. 1992. Regional evolutions. *Brookings Papers on Economic Activity*: 23, 1–76.

Bleakley, Hoyt and Jeffrey Lin. 2012. Portage and path dependence. *Quarterly Journal of Economics* 127(2): 587–644.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2013. Does management matter? Evidence from India. *Quarterly Journal of Economics* 128(1): 1–51.

Buckley, Robert, Achilles Kallergis, and Laura Wainer. 2015. *The Housing Challenge: Avoiding the Ozymandias Syndrome*. New York: Rockefeller Foundation and New School.

Card, David, Jochen Kluve, and Andrea Weber. 2010. Active labour market policy evaluations: A meta-analysis. *Economic Journal* 120(548), F452-F477.

Carrillo, Paul E., Arun S. Malik and Yiseon Yoo. 2016. Driving Restrictions that Work? Quito's Pico y Placa Program. *Canadian Journal of Economics*, 49(4), 1536-1568.

Caselli, Francesco and Wilbur J. Coleman. 2001. The U.S. structural transformation and regional convergence: A reinterpretation. *Journal of Political Economy* 109(3): 584-616.

Chauvin, Juan Pablo, Ed Glaeser, Yueran Ma, and Kristina Tobio. 2017. What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States. *Journal of Urban Economics*, 98: 17-49.

Cingano, Federico and Fabiano Schivardi. 2004. Identifying the sources of local productivity growth. *Journal of the European Economic Association* 2(4): 720–742,

Combes, Pierre-Philippe, Gilles Duranton and Laurent Gobillon. 2016. The Costs of Agglomeration: House and Land Prices in French Cities. Mimeographed, University of Pennsylvania.

Combes, Pierre-Philippe, Gilles Duranton and Henry G. Overman. 2005. Agglomeration and the Adjustment of the Spatial Economy. Papers in Regional Science 84(3), 311-349.

Combes, Pierre-Philippe and Laurent Gobillon. 2015. The Empirics of Agglomeration Economies. In Gilles Duranton, Vernon Henderson, and William Strange (eds.) *Handbook of Regional and Urban Economics, volume 5A.* Amsterdam: Elsevier, 247–348.

Couture, Victor. 2015. Valuing the Consumption Benefits of Urban Density. Mimeographed, University of California Berkeley.

Cunha, Flavio, James J. Heckman, Lance Lochner, and Dimitriy V. Masterov. 2006. Interpreting the evidence on life cycle skill formation. In: Eric A. Hanushek and Finis Welch (eds): *Handbook of the Economics of Education, volume 1*: 697-812, Amsterdam: Elsevier.

Dasgupta, Basab, Somik V. Lall, and Nancy Lozano-Gracia. 2014. Urbanization and Housing Investment. World Bank Policy Research Working Paper No. 7110.

Davis, Donald R. and David E.Weinstein. 2002. Bones, bombs, and break points: the geography of economic activity. *American Economic Review* 92(5):1269-1289.

Davis, Donald R. and David E.Weinstein. 2008. A search for multiple equilibria in urban industrial structure. *Journal of Regional Science* 48(1): 29–65.

Davis, Lucas W. 2008. The Effect of Driving Restrictions on Air Quality in Mexico City. *Journal of Political Economy* 116(1): 38-81.

De Soto, Hernando, 2000. The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else. Basic Civitas Books.

Dixit, Avinash K. and Joseph E. Stiglitz. 1977. Monopolistic competition and optimum product diversity. *American Economic Review* 67(3), 297-308.

Donaldson, Dave. 2010. Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. *American Economic Review,* forthcoming.

Durand-Lasserve, Alain, Maÿlis Durand-Lasserve, and Harris Selod, 2015. *Land Delivery Systems in West African Cities: The Example of Bamako, Mali*. Washington, DC: World Bank Africa Development Forum.

Duranton, Gilles. 2016. Determinants of city growth in Colombia. *Papers in Regional Science*, 95(1), 101-132.

Duranton, Gilles and Erick Guerra. 2016. *Developing a Common Narrative on Urban Accessibility: An Urban Planning Perspective*. Washington DC: Brookings Institution.

Duranton, Gilles and Diego Martin. 2018. Entrepreneurship, specialisation, diversity, and sectoral growth in Colombian cities. Mimeographed in progress, Wharton School of the University of Pennsylvania.

Duranton, Gilles and Diego Puga. 2014. The growth of cities. In Philippe Aghion and Steven Durlauf (eds) *Handbook of Economic Growth, volume 2*, 781–853. Amsterdam: NorthHolland.

Duranton, Gilles and Diego Puga. 2015. Urban Land Use. In: G. Duranton, J. V. Henderson, and W. C. Strange, *Handbook of Regional and Urban Economics*, volume 5A. Amsterdam: Elsevier, pp. 467-560.

Faber, Ben. 2014. Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System. *Review of Economic Studies* 81(3): 1046–1070

Faggio, Giulia, and Henry Overman. 2014. The effect of public sector employment on local labour markets. *Journal of Urban Economics* 79: 91-107.

Farole, Thomas. 2011. Special Economic Zones in Africa: Comparing Performance and Learning from Global Experience, no. 2268 in World Bank Publications, The World Bank.

Farole, Thomas, & Akinci, Gokhan (Eds.). 2011. *Special Economic Zones: Progress, Emerging Challenges and Future Directions*. Washington, D.C. The World Bank.

Ferreyra, Maria Marta. 2017. Human capital in cities. Mimeographed, World Bank.

Field, Erica. 2007. Entitled to Work: Urban Property Rights and Labor Supply in Peru. *Quarterly Journal of Economics* 122(4): 1561-1602.

Foster, Vivien and Cecilia Briceno-Garmendia. 2010. *Africa's Infrastructure; A Time for Transformation*, World Bank, Washington DC.

Fujita, Masahisa, Paul Krugman and Anthony J. Venables. 1999. *The Spatial Economy; Cities, Regions and International Trade.* MIT Press, Cambridge MA

Galiani, Sebastian and Ernesto Schargrodsky. 2010. Property Rights for the Poor: Effects of Land Titling. *Journal of Public Economics* 94(9): 700-729.

Ganong, Peter and Daniel Shoag. 2017. Why has regional income convergence in the U.S. declined? *Journal of Urban Economics* 102, 76-90.

Gennaioli, Nicola, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2013. Human capital and regional development. *Quarterly Journal of Economics*, 128(1), 105–164.

Ghani, Ejaz , William R. Kerr, and Stephen O'Connell. 2014. Spatial determinants of entrepreneurship in India. *Regional Studies* 48(6), 1071-1089.

Ghani, Ejaz , William R. Kerr, and Ishani Tewari. 2013. Specialization, diversity, and Indian manufacturing growth. Mimeographed, Harvard Business School.

Glaeser, Edward L., and Joseph Gyourko. 2005. Urban decline and durable housing. *Journal of Political Economy* 113(2): 345-375.

Glaeser, Edward L., Heidi Kallal, José A. Scheinkman, and Andrei Shleifer. 1992. Growth in cities. *Journal of Political Economy* 100(6): 1126–1152.

Glaeser, Edward L., Sari Pekkala Kerr, and William R. Kerr. 2015. Entrepreneurship and urban growth: An empirical assessment with historical mines. *Review of Economics and Statistics* 97(2): 498-520.

Glaeser, Edward L., Giacomo A. M. Ponzetto, and Andrei Shleifer. 2016. Securing Property Rights. Mimeographed, Harvard University.

Glaeser, Edward L., and Albert Saiz. 2004. The rise of the skilled city. *Brookings-Wharton Papers on Urban Affairs* 5, 47–95.

Glaeser, Edward L., and Wento Xiong. 2017. Urban productivity in the developing world. *Oxford Review of Economic Policy*, 33(3) 373–404.

Gollin, Douglas, David Lagakos and Michael E Waugh, (2014) 'The Agricultural Productivity Gap', *Quarterly Journal of Economics* 129 (2) 939-93

Gollin, Douglas. 2014. "The Lewis Model: A 60-Year Retrospective." *Journal of Economic Perspectives*, 28(3): 71-88.

Gonzalez-Navarro, Marco and Climent Quintana-Domeque. 2016. Paving Streets for the Poor: Experimental Analysis of Infrastructure Effects. *Review of Economics and Statistics* 98(2): 254-267.

Grant, Matthew. 2017. Why special economic zones? Using trade policy to discriminate across importers. Mimeographed, Princeton University.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy* 118(3): 536-598.

Gyourko, Joseph and Raven Molloy. 2015. Regulation and Housing Supply. In Gilles Duranton, Vernon Henderson, and William Strange (eds.) *Handbook of Regional and Urban Economics, volume 5B*. Amsterdam: Elsevier, 1289-1338.

Harari, Mariaflavia and Maisy Wong. 2017. Long-term Impacts of Slum Upgrading: Evidence from the Kampung Improvement Program in Indonesia. Mimeographed, University of Pennsylvania.

Head, Keith and Thierry Mayer. 2014. Gravity Equations: Workhorse, Toolkit, and Cookbook. In Gopinath, G, E. Helpman and K. Rogoff (Eds): *Handbook of International Economics volume 4*. Amsterdam: Elsevier, 131-195.

Helpman, Elhanan. 1998. The Size of Regions. In: Pines, D., Sadka, E., Zilcha, I. (Eds.), *Topics in Public Economics: Theoretical and Applied Analysis*. Cambridge: Cambridge University Press, pp. 33–54.

Henderson Vernon, Tanner Regan, and Anthony J. Venables. 2017. Building the City: Urban Transition and Institutional Frictions. Mimeographed, London School of Economics.

Hornbeck, Richard. 2012. The enduring impact of the American Dust Bowl: Short-and long-run adjustments to environmental catastrophe. *American Economic Review* 102(4): 1477-1507.

Hornbeck, Richard, and Daniel Keniston. 2017. Creative destruction: Barriers to urban growth and the Great Boston Fire of 1872. *American Economic Review* 107(6): 1365-1398.

Hornbeck, Richard, and Suresh Naidu. 2014. When the levee breaks: black migration and economic development in the American South. *American Economic Review* 104(3): 963-990.

Hyman, Benjamin G. 2017. Can displaced labor be retrained? Evidence from quasi-random assignment to Trade Adjustment Assistance. Mimeographed, Wharton School of the University of Pennsylvania.

Iyer, Lakshmi, John D. Macomber, and Namrata Arora. 2011. Dharavi: Developing Asia's Largest Slum. Harvard Business School Case 710-004.

Kanemoto, Yoshitsugu. 1988. Hedonic prices and the benefits of public projects. *Econometrica* 56(4): 981-989.

Jensen, Robert. 2010. The (perceived) returns to education and the demand for schooling. *Quarterly Journal of Economics*, 125(2), 515-548.

Khandelwal, Amit and Matthieu Teachout. 2016. Special Economic Zones for Myanmar. Proecesed, International Growth Centre Policy Note, https://www.theigc.org/wp-content/uploads/2016/03/SEZs-in-Myanmar.pdf.

Kim, Sukkoo. 2008. Spatial inequality and economic development: Theories, facts, and policies. In Michael Spence, Patricia C. Annez, and Robert M. Buckley, eds., Urbanization and Growth. Washington DC: World Bank.

Kline, Patrick and Enrico Moretti. 2013a. Place based policies with unemployment. *American Economic Review Papers and Proceedings*, 103(3), 238-243.

Kline, Patrick, and Enrico Moretti. 2013b. Local economic development, agglomeration economies, and the big push: 100 years of evidence from the Tennessee Valley Authority. *Quarterly Journal of Economics* 129(1): 275-331.

Krugman, Paul R. 1991a. *Geography and trade*. Cambridge (Mass.): MIT Press.

Krugman, Paul. 1991b. History versus expectations. *Quarterly Journal of Economics* 106(2): 651-667.

Lipsey, Richard G. and Kevin Lancaster. 1956. The general theory of second best. *Review of Economic Studies* 24 (1): 11–32

Litman, Todd (2017) Generated Traffic and Induced Travel: Implications for Transport Planning. Working paper, Victoria Transport Policy Institute.

Matsuyama, Kiminori. 1991. Increasing returns, industrialization, and indeterminacy of equilibrium. *Quarterly Journal of Economics* 106(2): 617-650.

McKenzie, David, and Christopher Woodruff. 2014. What are we learning from business training and entrepreneurship evaluations around the developing world? *World Bank Research Observer* 29(1), 48-82.

Michaels, Guy, Ferdinand Rauch, and Stephen J. Redding. 2012. Urbanization and structural transformation. *Quarterly Journal of Economics* 127(2): 535–86.

Mihretu, Mamo and Gabriela Llobet, (2017) *Looking Beyond the Horizon: A Case Study of PVH's Commitment in Ethiopia's Hawassa Industrial Park*, World Bank.

Moretti, Enrico. 2004. Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics*, 121(1), 175–212.

Munshi, Kaivan and Mark Rosenzweig. 2016. Networks and misallocation: Insurance, migration, and the rural-urban wage gap. *American Economic Review* 106(1): 46-98.

Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny. 1989. Industrialization and the big push. *Journal of Political Economy* 97(5): 1003-1026.

Neumark, David and Helen Simpson. 2015. Place-based policies. In: G. Duranton, J. V. Henderson, and W. C. Strange, *Handbook of Regional and Urban Economics, volume 5B.* Amsterdam: Elsevier, pp. 1197-1287.

Parry, Ian W.H., Margaret Walls, and Winston Harrington. 2007. Automobile Externalities and Policies. *Journal of Economic Literature* 45(2): 373-399.

Quatraro, Francesco and Marco Vivarelli. 2015. Drivers of entrepreneurship and post-entry performance of newborn firms in developing countries. *World Bank Research Observer* 30(2), 277–305.

Redding, Stephen J., Daniel M. Sturm, and Nikolaus Wolf. 2011. History and industry location: evidence from German airports. *Review of Economics and Statistics* 93(3): 814-831.

Redding, Stephen J. and Matthew A. Turner. 2015. Transportation Costs and the Spatial Organization of Economic activity. In Gilles Duranton, Vernon Henderson, and William C. Strange (eds.) *Handbook of Regional and Urban Economics, volume 5B.* Amsterdam: Elsevier, 1339–1398.

Shapiro, Jesse M. 2006. Smart cities: Quality of life, productivity, and the growth effects of human capital. *Review of Economics and Statistics*, 88(2), 324–335.

Schoar, Antoinette. 2010. The divide between subsistence and transformational entrepreneurship. Innovation Policy and the Economy 10 (1), 57-81.

Shoup, Donald C. 2005. *The high cost of free parking.* Chicago: Planners Press.

Siodla, James. 2015. Razing San Francisco: The 1906 disaster as a natural experiment in urban redevelopment. *Journal of Urban Economics* 89: 48-61.

Small, Kenneth A. 2012. Valuation of Travel Time. *Economics of transportation* 1(1), 2-14.

Small, Kenneth A. and Erik T. Verhoef. 2007. *The Economics of Urban Transportation*. New York:Routledge.

World Bank. 2009. *World Development Report 2009: Reshaping economic geography.* Washington, D.C.: World Bank