

HUMAN CAPITAL EXTERNALITIES IN CITIES: IDENTIFICATION AND POLICY ISSUES

June 2004

Gilles DURANTON

London School of economics and CEPR

Houghton Street, London WC2A 2AE, UK

g.duranton@lse.ac.uk

Abstract. This chapter analyses a few key issues pertaining to human capital externalities in cities. It shows the difficulties of properly identifying such externalities and explores the implications of such externalities for policy prescriptions.

Keywords: human capital externalities, cities, identification, economic policy.

JEL numbers: H23, R19, R39

This paper is a working draft of a chapter written for eventual publication in the *Blackwell Companion to Urban Economics* edited by Richard Arnott and Daniel McMillen. I am very grateful to Adala Bwire for her thorough reading and constructive comments on this paper. Richard Arnott and an anonymous referee also provided some useful feedback.

1. Introduction

The case for corrective economic policies as stated in Econ101 is relatively straightforward. In some instances, there is a wedge between the *private* costs (or benefits) resulting from some choice made by an economic agent and the *social* costs (or benefits) accruing to society. Such a wedge occurs when markets fail to mediate properly some economic interactions, that is when there is an *externality*. In such a case, the privately optimal decision made by the agent does not lead to a socially optimal outcome. For instance, investors under-invest when they cannot appropriate all the positive returns from their investments. Firms over-produce goods whose production damages the environment at no cost to them, etc. Pigou's (1890) ingenious solution to this type of problem is to impose a tax (or a subsidy) so that the private and social costs are made equal. In other words, whenever there is an externality, an appropriately chosen Pigovian tax or subsidy can make agents *internalise* the external effects of their choices so that the privately optimal decision leads to a socially optimal outcome.

After exposing the details of this argument, the Econ101 textbook typically points out that such taxes and subsidies are difficult to implement empirically because of their informational requirements. Then the textbook usually ends the discussion on this topic and turns to something else. From a policy perspective however, this is where the real work ought to start. In this chapter, the objective is to show how in practice economists attempt to identify such externalities. With this in mind, this chapter will analyse the case of one set of externalities: those pertaining to the market for human capital and education. Human capital externalities are of particular interest to urban economics, as shall be made clear below.

The rest of this chapter is organised in the following way. Section 2 shows why human capital

externalities in cities matter. Section 3 outlines a simple theoretical model and shows how it has been estimated empirically. Section 4 highlights a first set of criticisms to the standard approach. Section 5 discusses crucial issues of model identification and shows how they matter for policy purpose. Section 6 describes alternative approaches for the estimation of human capital externalities in cities. Finally, section 7 concludes.

2. Why Study Human Capital Externalities? Why In An Urban Context?

Among all the externalities that economists have been thinking about, human capital externalities are 'special' in two respects. First, they are potentially of formidable importance for a number of reasons. Such externalities provide a strong justification for subsidies to education. If the private returns to education are only half the social returns, the optimal Pigovian policy is a 50% subsidy to education. If instead, the social returns are essentially equal to the private returns, no subsidy is needed. Given that most developed economies spend up to 10% of their income towards education and training (broadly construed), the numbers at stake are very large. It worth noting however that human capital externalities are important beyond the issue of the optimal split between public and private expenditure for education. As argued by Lucas (1988) and his followers, *human capital externalities could constitute the fundamental engine of growth and development*. If this were the case, governments could draw on education policy to speed up economic growth. Moreover, since Marshall (1890) human capital externalities are also accepted as one of the main reasons to justify the existence of cities. This is because human capital externalities may arise predominantly from direct (or face-to-face) interactions between people, which are themselves expected to be highly distance sensitive.

The second reason why human capital externalities are special is that they are particularly difficult to identify. To repeat, we speak of externality when market prices fail to reflect the true social costs and benefits of an action. Unfortunately most of the data collected around the world consist of simple measurable characteristics of economic agents or of recorded market transactions. Since prices and quantities only reveal private costs and benefits, *externalities, by their very nature, leave no obvious paper trail by which they can be tracked or measured.*

This being said, with some externalities, social costs and benefits may be measured indirectly without too many conceptual difficulties. For instance, the costs of urban congestion can be measured by counting how many people are stuck in traffic jams and see how much time is wasted there. The mechanism at play is relatively simple: by taking the road at peak hours, I slow down everybody else and I do not take this into account when entering my car. The social cost of traffic congestion is the cost of the time wasted by people stuck in traffic jams plus that of an increase in pollution. Admittedly, measuring these costs precisely is by no means an easy task but this is still feasible. Urban congestion is a reasonably well-circumscribed problem over which transport economists have made very significant advances.

Human capital externalities are much more problematic because the mechanism at play is far less obvious. As shown below, there are many mechanisms that can generate human capital externalities. These mechanisms call for different policy prescriptions. In some extreme cases, subsidising education may even be counterproductive. The fact that the social costs and benefits of education have many dimensions complicates the matter even more. Human capital externalities can be thought of having a positive effect on productivity and wages as well as criminal behaviour or even voting outcomes. In what follows, the discussion will be restricted to the effects of human capital externalities on wages and earnings.

To look at human capital externalities, the starting point of existing research is the following. Positive human capital externalities imply that *measures of aggregate human capital should matter in the determination of outcomes over and above individual characteristics*. In absence of experiments on the issue, there are two main avenues for research: cross-section or time-series analysis. Time-series analysis does not appear particularly appropriate. Isolating the effects of an increase in education on aggregate output and confronting this to the private returns to education may be a hopeless task because of the incredibly large number of confounding factors that may affect changes in aggregate output over time. Instead, most research focuses on cross-section analysis conducted either at the cross-country or at the sub-national level. Cross-country analysis is made very difficult by the large number of institutional factors that may affect the outcomes of different countries. The second major problem is that comparing education data across countries is also very difficult. Sub-national analysis seems 'easier' to conduct because in many countries labour market data sets are of good quality and have become widely available. Such data typically records individual wages, education, and location for large samples of workers across a given country.

Among existing sub-national units, cities are of particular interests for two reasons. First, as highlighted above, human capital externalities may be at the root of the existence of cities. These externalities are thus expected to manifest themselves strongly at this level of analysis. Second, urban areas when properly defined provide economically meaningful economic units of analysis as opposed to arbitrarily defined administrative regions or states. Note however that by conducting the analysis across cities we give up on any attempt to measure country-wide human capital externalities. This may not be a serious issue when looking at wages or crime (which may be to a large extent determined locally) but potentially more problematic

when interested in voting behaviour.

3. The Standard Approach To The Analysis Of Human Capital Externalities In Cities

Consider an economy with workers (sub-indexed i or j) living in cities (sub-indexed a). The *social* output of worker i with human capital h_i and living in city a is given by:

$$y_i = (A + B_a)h_i . \tag{1}$$

where A is a technological parameter independent of location and B_a is a city specific parameter. At the same time, the earnings of this worker are:

$$w_i = Ah_i + D_a . \tag{2}$$

A straightforward comparison of (1) and (2) shows that workers do not receive the full value of their social product. Worker i cannot appropriate the part of her social output given by $B_a h_i$.

At the same time however, this worker benefits from being in city a and receives D_a as part of her earnings. This latter quantity will receive different theoretical interpretations in what follows. For the time being, it is convenient to think of it as the part of the external output of the other workers in the same city that accrues to worker i . Put differently, *there is a reciprocal externality within cities: workers do not receive the full surplus they create but instead receive part of the surplus created by the others.*

Note that the specification of equations (1) and (2) is additive rather than multiplicative. This is

mostly for simplicity and does not matter here (it does however in empirical work where multiplicative specifications are preferred because they fit the data better). It is also useful to bear in mind that empirically we can only observe the wage w_i , some proxy for human capital h_i and a few aggregates variables relating to city a but not the social output y_i .

To finish the description of the model, assume that the cost of human capital h_i is:

$$C_i = c_i h_i^\alpha, \tag{3}$$

with $\alpha > 1$. The cost shifter, c_i , can vary across individuals to reflect their differences in intrinsic abilities. Note that in a static context this cost may be interpreted as both a cost of acquisition and maintenance of human capital.

At the free-market equilibrium, the investment in human capital of worker i is chosen so as to maximise $w_i - C_i$. After simplification, we find equilibrium investment to be equal to:

$$h_i = \left(\frac{A}{\alpha c_i} \right)^{1/(\alpha-1)}. \tag{4}$$

The socially optimal human capital investment is instead such that it maximises $y_i - C_i$. It is given by:

$$h_i^* = \left(\frac{A+B}{\alpha c_i} \right)^{1/(\alpha-1)}. \tag{5}$$

The optimal Pigovian tax is to subsidise the returns to human capital in equation (2) by offering worker i an amount B_a per unit of human capital. Equivalently, one may subsidise the cost of acquiring human capital by a fraction $B/(A+B)$ so that the worker faces a cost of acquiring human capital equal to $C_i = Ac_i h_i^\alpha / (A+B)$ rather than (3).

What are human capital externalities about in this model? The standard story about human capital externalities in cities is eloquently summarised by Lucas (1988): "Most of what we know we learn from other people. We pay tuition to only a few of these teachers, either directly or indirectly by accepting lower pay so we can hang around them, but most of it we get for free, and often in ways that are mutual – without a distinction between student and teacher". Indeed, to write this paper I have read and benefited freely from a nice survey on human capital externalities in cities by Enrico Moretti (2004) from whom I borrowed the quote above. To write this chapter, I also built on previous work conducted with Sylvie Charlot (Charlot and Duranton, 2003). Again, she did not receive any direct compensation despite contributing to this chapter indirectly. Arguably such external effects take place across the board, in many industries and not only academia.

To be more precise about human capital externalities in an urban context, assume that worker i 's human capital directly benefits N other workers in the city by an amount $b h_i$. This 'interaction group' of N workers with whom worker i interacts is assumed to be a representative sample of workers in city a . At the same time, worker i also benefits from the human capital investment made by all other workers in the interaction group. With our notations, summing across all workers j part of the interaction group of worker i this implies:

$B_a = bN$ and $D_a = \sum b h_j = bN\bar{h}_a$, where \bar{h}_a is the average human capital in city a and N is the (unknown) size of the interaction group. Equation (2) can thus be re-written in the following manner

$$w_i = Ah_i + bN\bar{h}_a, \quad (6)$$

This equation (as well as many closely related specifications assuming different functional forms) can be estimated by means of regression analysis. The data needed for this exercise must be at the individual level. This data must contain the wage of each worker, a set of human capital characteristics (such as schooling but also labour market experience, etc), and possibly further individual controls. To estimate equation (6), a set of aggregate (i.e., city level) characteristics is also needed. Average schooling (or the fraction of university graduates) in the city is of course of particular interest here. The size of the interaction group may be taken as constant across cities or may be expected to increase with city population. In this case, the coefficient on city population can also be informative about the extent of human capital externalities.

Jim Rauch in 1993 was the first to estimate a specification with the same flavour as (6). He found that a one year increase in average schooling in a city would raise the expected wage of any of its workers whose education is unchanged by 3 to 5%. Conversely, increasing a worker's education by one year and keeping average city schooling constant raises her wage by around 5%. He found these estimates to be robust to the inclusion of a large number of individual and city controls. Stated differently, *Rauch's key empirical result is that external returns to education are roughly of the same order of magnitude as private returns (3 to 5% as opposed to 5%)*. Given that Rauch conducted his analysis very carefully and used a very large sample of

workers, this correlation between wages and average human capital was unlikely to be accidental. It attracted a large amount of attention and was successfully replicated many times.

If we accept this finding at face value and assume a causal relationship as in the model, the policy implications are the following: large subsidies to education are needed across the board because strong human capital externalities induce workers to grossly under-invest in their human capital. Related to this is the fact that education everywhere is heavily subsidised. In this respect, note that human capital externalities are not the only market failure mechanism justifying subsidies to education. Credit-constraints, lack of insurance, and sub-optimal parental investments are three other important motivations for public intervention in education. These other motives are justified by market failures surrounding the acquisition of human capital in equation (3). In this chapter, the focus is instead on equations (1) and (2).

So much for average city schooling, \bar{h}_a , what about the size of the interaction group, N ? Another robust finding in the literature is that the coefficient on city population N_a (used to proxy for the size of the interaction group) in the estimation of (6) is usually positive and significant. This finding has further important implications for urban policy (see the chapter by Rosenthal and Strange in this volume for more details on this).

4. Some Possible Problems With The Standard Approach

After making sure that an empirical finding is not a coincidence by examining slight changes in the specification, it is useful to envision two kinds of criticisms. They are both potentially very serious but each of a different nature. In this section, we will be concerned by problems that could affect the magnitude of the coefficients within the proposed theoretical framework.

Practically this means addressing a range of relatively narrow concerns about equation (6). The second series of problems regard deeper issues about model selection, i.e., whether or not we can interpret the empirical results in the light of the proposed model. These identification/specification concerns will be examined in the next section where we will consider whether the estimation of (6) vindicates the theoretical model laid down in equations (1) and (2).

There are many potentially confounding factors that may introduce a spurious correlation between aggregate human capital and individual wages. First, it is worth noting that *workers are not assigned exogenously to cities and that some characteristics of cities and workers remain unobserved*. For instance, two observationally equivalent workers (sex, education, age, etc) may have nonetheless very different abilities, ambition and dedication to their work. In a systematic way, these two workers may have a tendency to choose to live and work in very different cities. Furthermore, cities experience shocks on their local economies which are very difficult to observe but nonetheless can have a simultaneous impact on wages and the skill-composition of the workforce. For instance, it can be argued that the internet boom in the late 1990s led to both higher wages and an influx of educated workers in the San Jose area.

Allowing for worker mobility, it is also natural to ask whether causality in equation (6) is running from high average schooling to high wages or instead from high wages to high schooling. Furthermore, cities with high wages (after controlling for individual characteristics) may provide more education and thus enjoy more educated workers. For instance the revenue derived from a local natural resource may be used to increase local education. Alternatively, 'high-wage' cities may attract primarily more educated workers who will be able to benefit more from them. In these two cases, causality is running from high-wages to high levels of

human capital and not the other way round. Put differently, when any of these two arguments applies and is not dealt with the coefficient on average education will overstate the true extent of human capital externalities.

At the same time however, cities with good unobservable amenities may attract large numbers of highly educated workers who value these amenities more. For instance a city with a milder climate may appeal primarily to highly educated workers who put a higher premium on such amenity. In such a city, the price of land will be higher because each worker may be willing to consume more land (e.g., to have a larger garden) and because more (highly educated) workers will be willing to live in this city. Higher land prices will in turn reduce the demand for land from firms. If labour and land are imperfect substitutes in the production function, higher land prices will imply a lower productivity of labour and in turn lower wages. Put differently, when amenities like this are not included in the regression the coefficient on average education will understate the true extent of human capital externalities.

To summarise, *average schooling may be suspected of being determined simultaneously with wages*. Depending on which one of the stories above applies, the coefficient on average schooling may be over-estimated (as in the high-wage cities attracting high-workers story) or under-estimated (in the unobserved amenity argument). This type of concern has been taken very seriously in the literature. To deal with it, instrumental variables have been considered.

An instrumental variable is a variable that is correlated with the explanatory variable suspected of being endogenous but uncorrelated with the residual of the main regression. The procedure works as follows. In the 'instrumental regression', the variable suspected of endogeneity is regressed on the set of instrumental variables and a predicted value is computed. This predicted

value (which is determined only by a 'truly' exogenous set of instrumental variables) is then used as a regressor in the main regression. The idea behind this procedure is to re-create a *ceteris paribus* whereby the effects of exogenous changes in the dependent variables can be properly assessed. In our case, a good instrument for average schooling would affect the schooling of the majority of workers in a given location without being *otherwise* correlated with local wages.

Acemoglu and Angrist (2000) argue that differences in school compulsory attendance laws and child labour laws in US states over the 20th century provide good instruments for average schooling, at least for secondary education at the state level. In their preferred estimation, they obtain very small external returns to education. In the same spirit but using different instruments for average schooling (lagged demographic variables and the presence of land grant colleges), Moretti (2000) shows that the share of university graduates in a city has a strong effect on individual earnings. These results are not as contradictory as they seem. Acemoglu and Angrist (2000)'s results regard mostly the supply of secondary education whereas Moretti (2000)'s regression contains the share of university graduates. It may well be that human capital externalities are mostly generated by university graduates rather than workers who were forced to stay in high-school up to the age of 15 or 16 by the legislation. It also appears that the unit of analysis (US states in Acemoglu and Angrist versus metropolitan areas in Moretti) matters a lot.

To deal with workers unobserved heterogeneity, longitudinal data are desirable. The idea here is that there may be permanent but unobserved worker characteristics that play an important part in the story. Such unobserved characteristics are ambition, motivation, trustworthiness, self-discipline, etc. These characteristics are usually unobserved by the statistician but play a

very important role in the labour market. If workers sort themselves according to their characteristics in a manner that is correlated with high average schooling, the coefficient on this latter variable will be biased. The effects of these permanent characteristics for a given worker is called a fixed-effect. When workers are observed at least twice in the data at different periods, it is possible to estimate their fixed-effects. What is needed to do this estimation is some mobility of workers across cities. Otherwise, if no worker ever moves, it is impossible to distinguish what is caused by the permanent characteristics of workers from the effect of the permanent characteristics of their cities. In short, by using a panel of workers we can condition out all the permanent characteristics of workers. However, note that using a panel of workers is not as fool-proof as is sometimes claimed. Imagine for instance that the workers with high fixed-effects tend to move to cities receiving a positive shock (e.g., more ambitious workers flocked to Silicon Valley at the beginning of the internet boom). This generates a correlation between the true residual in the regression and individual fixed-effects that will bias the estimates. More generally, movers may not be a random sample of the population so that the estimation of the city effects is made with a biased sample of population.

Moretti (2000) develops a slightly different strategy and assumes a fixed-effect for each city/worker match. In this case, the source of identification is given by changes in the human capital composition of the city for 'constant' workers. Put differently, his estimation strategy allows him to estimate the impact on the wage of any worker of a change in aggregate human capital around this worker. Again, estimates may be biased since stayers may not be a random sample of the population. To get around this problem, Moretti (2000) tries a variety of different specifications. His estimates show that a 1% increase of university educated workers raise the average wage of workers whose education is unchanged by around 0.5%. Given that university educated workers have a wage about 50% higher than those of other workers, the external

returns to education implied by these estimates are again of the same order of magnitude as the private returns.

Although the evidence is not overwhelming, a few conclusions can be drawn at this stage. First there is a strong and robust correlation between most measures of average human capital and average wages after controlling for individual characteristics. This holds across most types of sub-national spatial units. Second, it also appears that at the city level, there is a causal effect of the share of university graduates on local wages. In other words, *there is some empirical support about causality running from average human capital to average wages in cities* as stated in equation (6).

5. Deeper Problems: Identification And Its Policy Implications

In the previous section, we mostly examined 'problems' that may complicate the estimation of equation (6) but without fundamentally questioning the assumptions made in (1) and (2). What we did was to assume that our basic story as told by equation (1) and (2) was true but consider that there could be some other economic phenomena that could come into play and create some *perturbations* in the basic estimates. It could also be the case that our estimation of equation (6) may be 'working' for reasons that have very little to do with equations (1) and (2). Put differently, it may be that we observe that average schooling in a city has a positive effect on average local wages but the true model may be unrelated to the existence of human capital externalities of the sort postulated so far. Put yet differently, *nothing that we have done so far guarantees us that we have tested 'the true model'*. In our case, this is particularly important from a policy perspective. This also matters a lot for our understanding of how cities tick.

Before exploring some alternatives to our basic model, let us insist more on this fundamental point. The 'Standard Paradigm' in applied economic analysis was first introduced by Alfred Marshall (1895) in his third edition of his *Principles of Economics*. His analogy was the following: predicting the exact movement of tides is very difficult because, although tidal movements are initially caused by gravity, they are also affected by many subtle meteorological factors. However, predicting the 'tendencies' can still be done by using gravity because meteorological factors only have a secondary influence on tides. As argued by Sutton (2000), from whom this argument is borrowed, some economic situations do indeed follow this analogy. Economists interested for instance in option pricing face a problem for which the variable to be explained (the price of options) depends on a small number of factors playing a systematic role. All these factors can be measured. The model explaining the price of options is 'self-evident' (although mathematically very difficult) since it is based on a strong arbitrage argument. Besides, the institutional rules that govern the actions of the participants are common knowledge. In this type of situation, economists are doing remarkably well at explaining the observed outcomes. To some extent, urban congestion externalities follow a similar pattern. With human capital externalities, unfortunately, model selection is a serious problem as there are many plausible stories aside from the one given above that could explain our findings so far. (As argued by Manski, 1993, this type of identification problem is pervasive in situations involving non-market interactions.)

5.1. Imperfect substitutability across workers

With respect to our model, the most obvious (and damaging) alternative is a simple story of supply and demand. Assume for simplicity that there are only two levels of human capital, high or low, super-indexed by H and L respectively. Let us take the human capital composition of

cities as given for now. The production function replacing (1) is now of aggregate nature.

Output in city a is now given by

$$Y_a = A(N_a^L)^\alpha (N_a^H)^{1-\alpha} . \quad (7)$$

where N_a^L is the number of low human capital workers in the city and N_a^H that of high human capital workers. In this model, there is no externality of any kind. Workers are paid at their marginal product, w_a^H and w_a^L . The average wage in the city is $\bar{w}_a = (N_a^L w_a^L + N_a^H w_a^H) / (N_a^H + N_a^L)$. Simple algebra then shows that for N_a^H sufficiently small:

$$\left. \frac{\partial \bar{w}_a}{\partial N_a^H} \right|_{N_a = Ct} > 0 . \quad (8)$$

According to (8) when high human capital workers are scarce an increase in their supply raises average wages. This is because workers with low levels of human capital gain more than high human capital workers lose. Put differently, keeping city population constant and assuming that high human capital workers are relatively scarce, this alternative model predicts that an increase in the proportion of high human capital workers increases the average wage in the city. Hence, even in absence of human capital externalities an increase in the supply of high human capital workers can raise the average wage in the city.

This competing explanation, put forward by Ciccone and Peri (2002) and Moretti (2000) would be particularly damaging for the whole idea of human capital externalities. *There may*

not be such thing as human capital externalities and all the regressions commented above would boil down to nothing more than bad estimations of production functions with imperfect substitutability. To deal with this issue, Moretti (2000) runs the same regressions as before but considers only highly educated workers. Finding that the wages of highly educated workers increases with their supply would indicate that positive human capital externalities more than offset the negative effects of relative supply. He finds strong support for this. However, Ciccone and Peri (2002) propose another estimation procedure where they look at the effects of an increase in average schooling keeping the relative supply of the different education groups constant. Using this 'constant composition' approach, they find only weak evidence of human capital externalities. To explain the divergence on this crucial point, further work is certainly needed.

5.2. Internalised externalities and ability externalities

Unlike the previous one, the two arguments explored in this sub-section do not deny the existence of some externalities related to human capital in cities but their policy implications differ radically from the standard prescriptions given about subsidising the acquisition of human capital.

First assume that (1), (2) and (3) hold but that the human capital externality, rather than taking place in a large group (i.e., the city), takes place in a much smaller unit (e.g., a firm) that for simplicity we model as a pair. In this story, workers are randomly paired *and then cooperatively invest* in the acquisition of their human capital. Put differently, we assume that for worker i paired with worker j , we have $D_a = b h_j$. In this case, the total income of the pair is

$$w_i + w_j = (A + b)h_i + (A + b)h_j. \quad (9)$$

In equilibrium i and j co-operatively maximise (9) with respect to their choice of human capital. Since $C_i = c_i h_i^\alpha$ and $C_j = c_j h_j^\alpha$, we get $h_i = ((A + b)/\alpha c_i)^{1/(\alpha-1)}$ and $h_j = ((A + b)/\alpha c_j)^{1/(\alpha-1)}$. It can then be verified that this equilibrium investment is optimal.

A discrepancy between social and observed private returns may be present when the returns to investment are delayed (like in law firms or consultancies where it takes a long time to become partner). Of course, this story may not appear very plausible with basic formal schooling. Typically schooling choices are made well before workers are paired (e.g., before they join a firm). However, when it comes to training, it may be highly relevant and explain why firms heavily subsidise the training of their workers.

Another possible argument is that the externality may not be caused by human capital but instead by the innate abilities of workers. In our framework, it would mean that the externality is not in h_i but instead in $1/c_i$ (to the extent that a low c_i captures high abilities). The individual production function would then be:

$$y_i = A h_i + b N/c_i, \quad (10)$$

and within the interaction group, the term D_a would be of the form: $D_a = \sum b/c_j$. In this case of ability-driven externality, the investment in human capital by workers is optimal. With human capital investment given by (4), it is quite clear that in equation (2) there will be an apparent correlation between average human capital and individual wages. *This correlation is however spurious here since both individual wages and average human capital are determined*

by an ability externality. Since abilities (captured by c_i) are usually not known to the analyst, we are facing a case of a missing variable bias.

A similar problem plagues the interpretation of the coefficient on individual education. Does the effect of individual schooling on wages really reflect the returns to human capital acquisition or does it only reflect innate abilities acknowledging that more able individuals get better degrees? Labour economics has been wrestling with this problem for a long time. The interested reader may refer to Card (1999) for a survey. However, the conclusion of Card and much of the literature on the topic is that the measured returns to educational attainment indeed largely reflect the returns to schooling and not those to ability. We may suspect the same applies to human capital externalities. The suspicion is strong but this is still un-proven.

5.3. Market failures 'elsewhere'

We have explored three objections so far (another complementarity, internalised externalities, and missing variable). A fourth type of objection may be made. It basically states that the model is not 'precise enough' about the real economic phenomena it attempts to capture. This matters a lot here because policy recommendations are sensitive to the exact details of the true model.

To understand this important point note first that, from an economic standpoint, cities are best viewed as a trade-off between 'agglomeration economies' and congestion costs. This argument is developed at length by Robert Helsley's chapter in this volume. These agglomeration economies (i.e., productive advantages coming from the spatial concentration of labour and capital) can be of different kinds. Alfred Marshall (1890) distinguished between the benefits

from labour pooling (a large market gives firms a constant pool of skilled workers), those from input/output linkages (a large market favours the existence of specialised suppliers) and a better circulation of ideas (the famous "industry is in the air" argument). Only this last motive is directly consistent with the arguments given in Section 3 to motivate equations (1) and (2). However, all three explanation may link directly with human capital. Put differently, it is now time to enter the micro-economic structure of equations (1) and (2) and attempt to *understand precisely how human capital externalities percolate*.

When trying to reconcile Marshall's first two arguments for agglomeration in a human capital framework, it may be argued for instance that having a greater number of skilled workers may facilitate matching between employers and employees. This is because a larger market for skills will make it easier for firms to find employees with the exact characteristics they require. At the same time, a large labour market will make it easier for highly skilled employees to find a job corresponding to their skills. In a different vein, with a larger supply of human capital in a city we expect an increase in the number of specialised suppliers. A large range of specialised suppliers may again have productive benefits and raise the marginal product of all workers in the city. In short, *most mechanisms generating local increasing returns to scale can be enriched to take human capital into account and generate external effects of human capital*. Hence equations (1) and (2) need not rely on direct interactions across workers.

The external effects that appear in the matching and input/output linkages argument that are sketched in the previous paragraph are nonetheless of a nature different from those in the standard story of Section 3 where direct interactions between workers are put forward. These alternatives may also require very different policy instruments. For instance, if the observed correlations are mostly explained by a thick local labour market argument, it is very unclear

whether the best thing to do is to encourage the acquisition of human capital or instead to attempt to reduce matching frictions so that workers are paid at their marginal product.

This point is further illustrated by the following example. Imagine that cities with larger shares of university graduates generate more employment in local R&D. Then assume that the successful innovations of the R&D sector cannot be fully appropriated by the inventors and benefit other firms in the same city. What is the best policy prescription: encourage the acquisition of human capital or strengthen intellectual property rights? This is far from obvious. Here again, the answer will depend on the precise microeconomic foundations of the model.

6. Further Attempts at Identification

To deal with these problems, the way forward is to identify more precisely how these externalities work. This is a very hard task because as already argued, externalities tend to leave no paper trail. *All the approaches explored so far were only making some inference about their existence without ever being able to observe them directly.* Only a very small number of papers have attempted to identify human capital externalities. Two main methods have been used: clever use of existing 'standard' data and use of more 'exotic' data.

Patent data have been used by economists for a long time to measure innovative output. Jaffe, Trajtenberg and Henderson (1993) start by re-asserting that patents can also be viewed as inputs in the innovative process. Innovators when filling an application for a patent must cite which other patented innovations they used in their own work. In a nutshell patents are like academic papers: they must contain a 'bibliography'. The crucial difference with academic

papers where authors may cite liberally (and arbitrarily) is that for patents the bibliography is not done by the author but by the equivalent of an independent editor, the patent officer. For this reason, patent citations may contain valuable information about the real intellectual sources of innovations.

The null hypothesis of Jaffe et al. (1993) is that the citation trail is neutral with respect to geography. In absence of 'knowledge spill-overs' (which may be thought of as the counterpart to human capital externalities when embodied in an innovation), we expect cited patents to be distributed like any 'control group' of patents. In other words, if California accounts for 20% of past patents in micro-electronics in the US, we expect any new patent in this industry to quote 20% of Californian patents regardless of where the new innovator is located. The key result of Jaffe et al. (1993) is that this null hypothesis is strongly rejected at three different spatial scales: US vs. rest of the world, between US states and between US metropolitan areas. US innovators quote patents from the same country, state and city much more than patents from other countries, states and cities even after controlling for the spatial distribution of the research output across sectors.

Taking a different direction, Charlot and Duranton (2003) exploit a unique survey recording communication practices for around 6,000 French workers in 1997 (with information about whom they communicate with and how). Their identifying assumption is very much in line with the standard story told by Lucas (1988) about human capital externalities. More educated (and more populated) cities should favour communication and in turn more communication should have a positive impact on individual earnings. The strength of the 'communication' externalities can then be computed as the effects of average urban schooling (and city population) on communication multiplied by the effect of communication on earnings.

Consistent with these two hypotheses, Charlot and Duranton (2003) find that workplace communication is positively associated with earnings. Furthermore, average urban schooling (and city population) is positively associated with workplace communication. Between a tenth and a quarter of the effects of average urban schooling (and city population) appear to permeate through communication externalities.

7. Conclusions

Given how little we know and the large number of possible channels through which human capital externalities may percolate, it is far too early to make definitive conclusions. Note however that the last 10 years have seen tremendous progress on the subject. The first serious paper documenting the issue was Rauch in 1993. Over the last ten years, much has been learnt about human capital externalities in cities.

If we go back to the general argument about 'how to do economics' that is raised above, it is worth noting that human capital externalities is one of these subjects where problems of model selection are very important. Progress on this type of issue is always going to be slow and tentative. When I first spoke with one of the editors (Richard Arnott) about this chapter, he encouraged me to write on this topic but to highlight the difficulties of the topic he wrote to me: "There seems to be something special about large cities that make them centres of culture, art, and innovation. Cities attract not just the bright and knowledgeable, but also the bold, unconventional, innovative, and ambitious. Cities are where you go to make a name, and cities are where you find access to patrons and venture capital." Capturing this type of argument theoretically through neat little models and then estimating them is certainly not an easy task.

To summarise, after ten years of work on human capital externalities in cities, there is a strong suspicion that 'something is going on' here, a weaker suspicion that such externalities may be quite large and a hunch that direct interactions may not be everything. We are still quite far from policy recommendations strongly grounded in evidence. This should remain the final goal however. From a narrow perspective, it may be argued that if only 10% of resources in education were misallocated (arguably a very conservative estimate), the number would still be huge: \$bn 70 for the US each year would be wasted!! This probably justifies a fair amount of research resources devoted to the topic. From a broader perspective, a successful education is clearly a life-changing experience (and often for the best, academics like to think) which certainly impacts on others around. The proper provision of education is obviously crucial, at least in this respect.

7320 words in total

References

- Acemoglu, D., Angrist, J. (2000) How large are human capital externalities? Evidence from compulsory schooling laws. *NBER Macroeconomics Annual* 15(0), 9-59.
- Card, D. (1999) The causal effect of education on earnings. In: Ashenfelter, O. and Card, D. (eds) *Handbook of Labor Economics* 3. North-Holland, Amsterdam.
- Charlot, S., Duranton, G. (2003) Communication externalities in cities. Unpublished manuscript, London School of Economics.
- Ciccone, A., Peri G. (2002) Identifying human capital externalities: Theory with an application to US cities. CEPR discussion paper 3350.
- Jaffe, A. B., Trajtenberg, M., Henderson, R. (1993) Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics* 108(3), 577-598.
- Lucas, R. E. Jr, (1988) On the mechanics of economic development. *Journal of Monetary Economics* 22(1), 3-42.
- Manski, C. F. (1993) Identification of endogenous social effects: The reflection problem. *Review of Economic Studies* 60(3), 531-542.
- Marshall, A. (1890, 1895) *Principles of Economics*. Macmillan, London (first and third editions).
- Moretti, E. (2000) Estimating the social return to education: Evidence from longitudinal and cross-sectional data. Unpublished manuscript, UCLA.
- Moretti, E. (2004) Human capital externalities in cities. In Henderson, J. V., and Thisse, J.-F. (eds.) *Handbook of Regional and Urban Economics* 4. North Holland, Amsterdam.
- Pigou, A. C. (1920). *The Economics of Welfare*. Macmillan, London.
- Rauch, J. E. (1993) Productivity gains from geographic concentration of human capital:

evidence from the cities. *Journal of Urban Economics* 34 (3), 380-400.

Sutton, J. (2000) *Marshall's Tendencies: What Can Economists Know?* MIT Press, Cambridge, MA.

Note for further readings

This chapter is at the cross-road between several threads of literatures. Here are some leads.

- On human capital externalities, the next step is certainly the survey by Moretti (2004) in the *Handbook of Regional and Urban Economics*. This survey is much more technical than the present chapter. It discusses in depth the material in Sections 4 and 5.1.

- On the empirics of agglomeration economies more generally, the chapter by Rosenthal and Strange in this Volume should be an obvious point of departure. Their more involved survey in the *Handbook of Regional and Urban Economics* is certainly the next step.

- On the theories of agglomeration, the chapter by Helsley in this volume is the place to start. The chapter by Duranton and Puga in the *Handbook of Regional and Urban Economics* is again a natural second step.