HAL
open science

The Empirics of Agglomeration Economies

Pierre-Philippe Combes, Laurent Gobillon

To cite this version:

Pierre-Philippe Combes, Laurent Gobillon. The Empirics of Agglomeration Economies. 2014. halshs-01071761

HAL Id: halshs-01071761

https://shs.hal.science/halshs-01071761v1

Preprint submitted on 6 Oct 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PARIS SCHOOL OF ECONOMICS
ÉCOLE D'ÉCONOMIE DE PARIS

WORKING PAPER N° 2014 – 31

The Empirics of Agglomeration Economies

Pierre-Philippe Combes
Laurent Gobillon

JEL Codes: R12, R23, J31

Keywords: Agglomeration gains, Density, Sorting, Learning, Location choices

PARIS-JOURDAN SCIENCES ECONOMIQUES
48, BD JOURDAN – E.N.S. – 75014 PARIS
TÉL. : 33(0) 1 43 13 63 00 – FAX : 33 (0) 1 43 13 63 10
www.pse.ens.fr

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE – ECOLE DES HAUTES ETUDES EN SCIENCES SOCIALES
ÉCOLE DES PONTS PARISTECH – ECOLE NORMALE SUPÉRIEURE – INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQU

The Empirics of Agglomeration Economies

Written for the Handbook of Urban and Regional Economics, vol. 5
Eds. Gilles Duranton, Vernon Henderson, Will Strange, Elsevier 2015.

Pierre-Philippe Combes[ab]
Aix-Marseille School of Economics and *Sciences Po Paris*

Laurent Gobillon[ac]
INED - Paris School of Economics

September 17, 2014

Abstract

We propose an integrated framework to discuss the empirical literature on the local determinants of agglomeration effects. We start by presenting the theoretical mechanisms that ground individual and aggregate empirical specifications. We gradually introduce static effects, dynamic effects, and workers' endogenous location choices. We emphasise the impact of local density on productivity but we also consider many other local determinants supported by theory. Empirical issues are then addressed. Most important concerns are about endogeneity at the local and individual levels, the choice of a productivity measure between wage and TFP, and the roles of spatial scale, firms' characteristics, and functional forms. Estimated impacts of local determinants of productivity, employment, and firms' locations choices are surveyed for both developed and developing economies. We finally provide a discussion of attempts to identify and quantify specific agglomeration mechanisms.

Keywords: Agglomeration gains, density, sorting, learning, location choices.

JEL classification: R12, R23, J31

[a]This paper was prepared for the Handbook of Urban and Regional Economics, vol. 5, Gilles Duranton, Vernon Henderson, Will Strange (eds.), to be published by Elsevier, 2015. We are grateful to Gilles Duranton, Vernon Henderson, Jeffrey Lin, Steve Ross, and William Strange, as well as participants at the Handbook conference at the Wharton School of the University of Philadelphia for useful comments and discussion.

[b]Aix-Marseille University (Aix-Marseille School of Economics), CNRS, & EHESS, 2 Rue de la Charité, 13002 Marseille, France and Sciences Po, Economics Department, 28, Rue des Saints-Pères, 75007 Paris, France (e-mail: ppcombes@gmail.com; website: http://www.vcharite.univ-mrs.fr/pp/combes/). Also affiliated with the Centre for Economic Policy Research (CEPR).

[c]Institut National d'Etudes Démographiques, 133 Boulevard Davout, 75980 Paris cedex 20, France (e-mail: laurent.gobillon@ined.fr; website: http://laurent.gobillon.free.fr/). Also affiliated with the Paris School of Economics, the Centre for Economic Policy Research (CEPR), and the Institute for the Study of Labor (IZA).

Contents

1 Introduction **2**

2 Mechanisms and corresponding specifications **5**
2.1 Static agglomeration effects and individual skills . . . 5
2.1.1 Separate identification of skills and local effects . . . 5
2.1.2 Heterogeneous impact of local effects . . . 13
2.2 Dynamic impact of agglomeration economies . . . 15
2.3 Extending the model to local worker-firm matching effects . . . 19
2.4 Endogenous inter-temporal location choices . . . 21

3 Local determinants of agglomeration effects **23**
3.1 Density, size and spatial extent of agglomeration effects . . . 24
3.2 Industrial specialisation and diversity . . . 27
3.3 Human capital externalities . . . 31

4 Estimation strategy **34**
4.1 Wages versus TFP . . . 35
4.2 Endogeneity issues . . . 37
4.3 Dealing with endogenous local determinants . . . 38
4.3.1 Local fixed effects . . . 39
4.3.2 Instrumentation with historical and geological variables . . . 39
4.3.3 Generalised Method of Moments . . . 41
4.3.4 Natural experiments . . . 43
4.4 Tackling the role of firm characteristics . . . 44
4.5 Other empirical issues . . . 46
4.5.1 Spatial scale . . . 46
4.5.2 Measures of observed skills . . . 47
4.5.3 Functional form and decreasing returns to agglomeration . . . 48
4.5.4 Spatial lag models . . . 49

5 Magnitudes for the effects of local determinants of productivity **50**
5.1 Economies of density . . . 51
5.2 Heterogeneous effects . . . 55
5.3 Spatial extent of density effects . . . 58
5.4 Market access effect evaluated using natural experiments . . . 59
5.5 Specialisation and diversity . . . 61
5.6 Human capital externalities . . . 62
5.7 Developing economies . . . 63

6 Effects of agglomeration economies on outcomes other than productivity **66**
6.1 Industrial employment . . . 66
6.1.1 From productivity externalities to employment growth . . . 66
6.1.2 Total employment, specialisation, diversity, and human capital . . . 71
6.1.3 Dynamic specifications . . . 72
6.2 Firms' location choices . . . 73
6.2.1 Strategies and methodological concerns . . . 74
6.2.2 Discrete location choice models . . . 75
6.2.3 Firm creation and entrepreneurship . . . 78

7 Identification of agglomeration mechanisms **79**
7.1 Labour mobility, specialisation, matching, and training . . . 80
7.2 Industrial spatial concentration and co-agglomeration . . . 82
7.3 Case studies . . . 87

8 Conclusion **89**

1 Introduction

Ongoing urbanisation is sometimes interpreted as an evidence of gains from agglomeration that dominate its costs, otherwise firms and workers would remain sparsely distributed. One can imagine, however, that the magnitude of agglomeration economies depends on the type of workers and industries, as well as on the period and country. This is a first motivation to quantify agglomeration economies precisely, which is the general purpose of the literature reviewed in this article. Moreover, firms' and workers' objectives, profit and utility, are usually not in line with collective welfare or the objective that some policy makers may have in particular for productivity or employment. Even if objectives were identical, individual decisions may not lead to the collective optimum as firms and workers may not correctly estimate social gains from spatial concentration when they choose their location. Generally speaking, an accurate estimation of the magnitude of agglomeration economies is required when one tries to evaluate the need for larger or smaller cities. If one were to conclude that the current city size distribution is not optimal, such an evaluation is necessary for the design of policies (such as taxes or regulation) that should be implemented to influence agents' location choices towards the social optimum. Lastly, many a priori a-spatial questions can also be indirectly affected by the extent to which firms and workers relocate across cities, as for instance inequalities among individuals and the possible need for policies to correct them. Typically, inequality issues might be less severe when workers are mobile and they rapidly react to spatial differences in the returns to labour. Addressing such questions requires beforehand a correct assessment of the magnitude of agglomeration economies.

Agglomeration economies is a large concept that includes any effect that increases firms' and workers' income when the size of the local economy grows. The literature proposes various classifications for the different mechanisms behind agglomeration economies, from Marshall (1890) who divides agglomeration effects into technological spillovers, labour pooling, and intermediate input linkages, to the currently most used typology proposed by Duranton and Puga (2004) who rather consider sharing, matching, and learning effects. Sharing effects include the gains from a greater variety of inputs and industrial specialisation, the common use of local indivisible goods and facilities, and the pooling of risk; matching effects correspond to improvement of either the quality or the quantity of matches between firms and workers; learning effects involve the generation, diffusion, and accumulation of knowledge. Ultimately one would like an empirical assessment of the respective importance of each of these components. Unfortunately, the literature has not reached this goal yet and we will see that there are only rare attempts to distinguish the various channels behind agglomeration economies. They are mostly descriptive and we present them at the end of this article. We choose rather to detail the large literature that tries to evaluate the overall impact on local outcomes of spatial concentration, and of a number of other characteristics of the local economy such as its industrial structure, its labour force composition or its proximity to large locations. In other words, what is evaluated is the impact on some local outcomes of

local characteristics that shape agglomeration economies through a number of channels, not the channels themselves. Local productivity and wages have been the main focus of attention but we also present the literature that studies how employment and firm location decisions are influenced by local characteristics.

When estimating the overall impact of a local characteristic, such as the impact of local employment density on local productivity, it is not possible to know whether the estimated effect arises mostly from sharing, matching or learning mechanisms, or from all of them simultaneously. Most positive agglomeration effects can also turn negative above some city size threshold, or induce some companion negative effects, and one cannot say whether some positive effects are partly offset by negative ones, as only the total net impact is evaluated. Moreover, while some mechanisms imply immediate static gains from agglomeration, other effects are dynamic and influence local growth. We take into account all these theoretical issues in our framework of analysis, as this is required to correctly choose relevant empirical specifications, correctly interpret the results and discuss estimation issues. Crucially, even if the effects of mechanisms related to agglomeration economies are not identified separately, knowing for instance by how much productivity increases when one increases the number of employees per square meter in a city is crucial for the understanding of firms' and workers' location choices or for the design of economic policies.

We will see that the role of some local characteristics is already not that trivial to evaluate. Beyond some interpretation issues that we will detail, the main difficulty arises from the fact that one does not seek to identify correlations between local characteristics and a local outcome but causal impacts. Basic approaches can lead to biased estimates because of endogeneity concerns at both the local and individual levels. Endogeneity issues at the local level arise from either aggregate missing variables that influence both local outcomes and local characteristics, or reverse causality as better average local outcomes can attract more firms and workers in some locations, which in turn affects local characteristics. Endogeneity issues at the individual level occur when workers self-select across locations according to factors that cannot be controlled for in the specification, typically some unobserved abilities, or when they choose their location depending on the exact individual outcome that depends on individual shocks that can be related to local characteristics. Dealing with these various sources of endogeneity is probably the area where the literature has made the greatest progress over the last decade. It is not possible any more to evaluate the determinants of local outcomes without addressing possible endogeneity issues. Therefore we largely discuss the sources of endogeneity and the solutions proposed in the literature.

Since various agglomeration mechanisms are at work and the impact of many local characteristics on different local outcomes has been studied, it is necessary to first clarify the theories that are behind the specifications estimated in the literature. Section 2 starts from a simple model and the corresponding specification that emphasises the determinants of local productivity. This model is then progressively extended to encompass additional mechanisms, moving from static specifica-

tions to dynamic frameworks, while stressing the role of individual characteristics and individual location choices. This approach helps to clarify some of the endogeneity issues. Section 3 presents all the local characteristics whose impact on productivity is studied in the literature, and relates them to theory. With such a theoretical background in mind, we systematically discuss a series of empirical issues in Section 4, mostly endogeneity concerns at the local and individual levels, as well as the solutions proposed to tackle them. We also discuss the choice of a productivity measure between wage and TFP, and the roles of spatial scale, firms' characteristics, and functional forms. Magnitudes of estimated agglomeration effects on productivity are presented in Section 5, which covers in particular the effect of density, its spatial extent, and some possible heterogeneity of the impact across industries, skills, and city sizes. The section also presents the results of the most recent studies, which use a structural approach or exploit natural experiments, as well as results on the role of the industrial structure of the local economy (namely industrial specialisation and diversity) and human capital externalities. Recent results for developing economies are detailed separately as magnitudes are often not the same as for developed countries and their study is currently in expansion. In Section 6, estimated agglomeration effects on employment and firms' location choices instead of productivity are discussed, after starting with considerations related to theory and the choice of a relevant empirical specification. Finally, Section 7 presents attempts to identify the channels through which agglomeration economies operate. The identification of such channels is one of the current concerns in the literature.

The organisation of our article does not follow the development of the field over time. The literature started with the ambitious goal of estimating the impact of a large number of local determinants on employment growth at the city-industry level (Glaeser et al., 1992; Henderson et al., 1995). However, acknowledging some possibly serious interpretation and endogeneity concerns, the literature then became more parsimonious, focusing on static agglomeration effects on local productivity only (such as Ciccone and Hall, 1996; Glaeser and Maré, 2001; Combes et al., 2008a). This was also made possible thanks to the availability of new datasets with a panel dimension at the individual level. More recent contributions incorporate additional effects such as the dynamic ones already suggested in the previous literature (see De La Roca and Puga, 2012), or consider richer frameworks through structural models involving endogenous location choices and different sources of heterogeneity across firms and workers (such as Gould, 2007; Baum-Snow and Pavan, 2012). We make the choice to start with a simple but rigorous framework to analyse the effects of local determinants of productivity, which we then extend. Most of the contributions of the literature are ultimately encompassed, and this includes earlier ones focusing on employment growth. When referring to magnitudes of the effects, we focus more particularly on contributions later than those surveyed in Rosenthal and Strange (2004) but we refer to earlier contributions when they are relevant and useful.

Still, there are a number of related topics that we do not cover, mostly because they involve too

much material and the handbook editors made the choice of devoting separate chapters to them. In particular, a specific case where the effect of an agglomeration mechanism can be identified is technological spillovers and the links between agglomeration and innovation. This topic is covered by Carlino and Kerr (2015) who also discuss the literature on agglomeration and entrepreneurship, as it is often grounded on technological spillovers. Similarly, we do not cover the literature on the interactions between agglomeration economies and place-based policies, since it is considered in Neumark and Simpson (2015). Finally, we do not present the various attempts made to measure spatial concentration. Nevertheless, we refer to spatial concentration indices in the last part of the survey as some papers use them in regressions to attempt to identify mechanisms of agglomeration economies.

2 Mechanisms and corresponding specifications

It is not possible to discuss the estimation of agglomeration economies without first clarifying the theories and underlying mechanisms that are assessed empirically by the literature. This section presents these theories so that we can then correctly interpret estimates and discuss possible estimation issues.

2.1 Static agglomeration effects and individual skills

2.1.1 Separate identification of skills and local effects

Earlier literature studies agglomeration economies at an aggregate spatial level, the region or the city. An outcome in a local market is typically regressed on a vector of local variables. In this section, we focus mostly on the impact of the logarithm of density on the logarithm of workers' productivity, measured by nominal wage. This corresponds to the relationship considered by Ciccone and Hall (1996) who had a large impact on the recent evolution of the literature. The role of other local determinants such as market access, industrial diversity, or specialisation has also been considered, and will be detailed in Section 3. Other local outcomes such as industry employment growth or firms' location choices will be discussed in Section 6.

Let us first consider a setting without individual heterogeneity among firms and workers. Let $Y_{c,t}$ be the output of a representative firm located in market c at date t. The firm uses two inputs, labour $L_{c,t}$, and other factors of production $K_{c,t}$, such as land, capital or intermediate inputs. The profit of the firm is given by:

$$\pi_{c,t} = p_{c,t}Y_{c,t} - \omega_{c,t}L_{c,t} - r_{c,t}K_{c,t}, \tag{1}$$

where $p_{c,t}$ is the price of the good produced, $\omega_{c,t}$ is the wage rate on the local labour market, and $r_{c,t}$ is the unit cost of non-labour inputs. Suppose that the production function is Cobb-Douglas

and can be written as:

$$Y_{c,t} = \frac{A_{c,t}}{\alpha^{\alpha}(1-\alpha)^{1-\alpha}} (s_{c,t}L_{c,t})^{\alpha} K_{c,t}^{1-\alpha}, \tag{2}$$

where $0 < \alpha < 1$ is a parameter, $A_{c,t}$ is the local total factor productivity, and $s_{c,t}$ corresponds to local labour skills. As long as all local firms and workers are assumed to be identical, these quantities depend on c and t only. In turn, this is also the case for $p_{c,t}$, $w_{c,t}$, and $r_{c,t}$. Indeed, in a competitive equilibrium, an assumption we discuss below, the first-order conditions for the optimal use of inputs reduce to:

$$w_{c,t} = \left(p_{c,t} \frac{A_{c,t}}{(r_{c,t})^{1-\alpha}} \right)^{\frac{1}{\alpha}} s_{c,t} \equiv B_{c,t}\, s_{c,t}. \tag{3}$$

The local average nominal wage depends on labour skills, $s_{c,t}$, as well as on a composite local productivity effect, $B_{c,t}$. This equation is enough to encompass almost all agglomeration effects that the literature has considered. If one goes back as far as to Buchanan (1965), cities are places where firms and consumers share indivisible goods such as airports, universities, hospitals, which generate a first type of agglomeration economies. In that case, the composite labour productivity effect, $B_{c,t}$, and therefore the local average wage, are higher in larger cities because $A_{c,t}$ is larger due to the presence of local (public) goods. This corresponds to a first type of pure local externality in the sense that it is not mediated by the market. A second type of pure local externality, very different in nature, emerges when spatial concentration induces local knowledge spillovers that make firms more productive, as put forward in early endogenous growth models such as Lucas (1988). Again, this type of mechanism makes $A_{c,t}$ larger in larger cities. For the moment, we implicitly assume that all these effects are instantaneous and affect only current values of $A_{c,t}$. This is an important restriction that we discuss further below.

Economists have also emphasised a number of agglomeration mechanisms operating through local markets, sometimes referred to as 'pecuniary externalities'. Because access to markets is better in larger cities, the price of goods there, $p_{c,t}$, can be higher, and the costs of inputs $r_{c,t}$, lower. Both effects again make $B_{c,t}$ larger.[1] Ultimately, one would like to assess separately whether pure externalities or local market effects play the most significant role on local productivity, or whether, among market effects, local productivity gains arise from price effects mostly related to goods or inputs. However, such assessments are difficult and a large part of the empirical literature

[1] When firms sell to many markets, $p_{c,t}$ corresponds to the firm's average income per unit sold, which encompasses trade costs, and the present analysis can easily be extended, as shown by Combes (2011). Let $Y_{c,r,t}$ denote the firm's exports to any other market r. Output value is the sum of the value of sales on all markets, $p_{c,t}Y_{c,t} = \sum_r (p_{c,r,t} - \tau_{c,r,t})Y_{c,r,t} = \left(\sum_r (p_{c,r,t} - \tau_{c,r,t})\phi_{c,r,t}\right) Y_{c,t}$ where $p_{c,r,t}$ is the firm's price on market r, $\tau_{c,r,t}$ represents trade costs paid by the firm to sell on market r, and $\phi_{c,r,t} = \frac{Y_{c,r,t}}{Y_{c,t}}$ is its share of output that is sold there. As a result, $p_{c,t} = \sum_r (p_{c,r,t} - \tau_{c,r,t})\phi_{c,r,t}$ is the average of the firm's prices over all its markets net of trade costs and weighted by its share of sales on each market. The closer to large markets the firm is, the lower the trade costs and the higher this average price. Similarly when firms buy inputs from many markets, the closer these markets are, the lower the firms' average unit cost of inputs, $r_{c,t}$.

on agglomeration economies simply quantifies the overall impact on productivity of characteristics of the local economy. The previous discussion shows in particular that the positive correlation between wage and density can result from pure externalities as well as effects related to good or input prices.

Furthermore, city size generates not only agglomeration economies but also dispersion forces. Typically, the cost of inputs that are not perfectly mobile, $r_{c,t}$, with land at one extreme, is higher in larger cities. If competition is tough enough relative to the benefits from market access in large cities, the price of goods there $p_{c,t}$ can be lower than in smaller cities. Congestion on local public goods can also emerge, which reduces $A_{c,t}$. Note also that if local labour markets are not competitive, the right-hand side in Equation (3) should be multiplied by a coefficient that depends on the local bargaining power of workers. If workers have more bargaining power in larger cities, their nominal wages are higher, and this constitutes an agglomeration effect. Alternatively, a lower bargaining power in larger cities is a dispersion force. The correlation between wage and density only reflects the overall impact of both agglomeration economies and dispersion effects. While the net effect of spatial concentration is identified, it is not the case of the channels through which it operates. Conversely, if one wants to independently quantify the impact of market effects operating through $r_{c,t}$ and $p_{c,t}$, a strategy is required involving controls for pure externalities arising for instance from the presence of local public goods or local spillovers.

One can also consider the inclusion of controls for dispersion forces if data on local traffic congestion or housing/land prices for instance are available. This is a start to disentangling agglomeration economies and dispersion forces. Importantly, the motivation for introducing housing/land prices is their influence on the costs of inputs and not compensation for low or high wages in equilibrium such that workers are indifferent between places as in Roback (1982). Indeed, we are focusing here on the determinants of productivity and not on equilibrium relationships. Typically, land price is expected to have a negative impact on nominal wage in accordance with equation (3), while the equilibrium effect implies a positive correlation between the two variables. As wages and land prices are simultaneously determined in equilibrium, controlling for land or housing prices can lead to serious endogeneity biases difficult to deal with (see the discussion below in section 4). This suggests that, if land represents a small share of input costs, which is usually the case, it is probably better not to control for its price in regressions.

Testing the relevance of a wage compensation model and quantifying real wage inequalities between cities are interesting questions but they require considering simultaneously the roles of nominal wages, costs of living and amenities. These questions are addressed in a burgeoning literature in expansion (Albouy, 2009; Moretti, 2013) which we briefly discuss in conclusion. As far as the effect of agglomeration economies on productivity only is concerned, nominal wage constitutes the relevant dependent variable and there is no need to control for land prices as illustrated by our model.

Let us turn to the role of local labour skills, captured in Equation (3) by $s_{c,t}$. If workers have skills that are not affected by their location, typically inherited from their parents or acquired through education, one definitively does not want to include the effect of skills among agglomeration economies, since it corresponds to a pure composition effect of the local labour force and not an increase in productivity due to local interactions between workers. It is possible that, for reasons not related to agglomeration economies, higher skills are over-represented in cities. This can arise for instance if skilled workers value city amenities (related for instance to culture or night life) more than unskilled ones or if, historically, skilled people have located more in larger cities and transmit part of their skills to their children who stay there. If the estimation strategy does not control for the selection of higher skills in cities, other local variables such as density capture their role, and the impact of agglomeration economies can be overstated. Alternatively it is also possible that people are made more skilled by cities, through stronger learning effects in larger cities, or that skilled people generate more local externalities, as suggested by Lucas (1988). In that case, not controlling for the skill level in the city is the correct way for capturing the total agglomeration effect due to a larger city size. A priori, both the composition effect and the agglomeration effect can occur, and a local measure of skills or education can capture both. The aggregate approach at the city level discussed here does not consider individual heterogeneity and does not allow the separate identification of the two effects. This is its first important limit and an individual data approach is more useful for that purpose as detailed below.

Finally, a crucial issue is the time span of agglomeration effects. One can accept that productivity and then wages adjust very quickly to variations in market-mediated agglomeration effects (operating through changes in $r_{c,t}$ and $p_{c,t}$), but they definitely do not for variations of most pure local externalities that can affect $A_{c,t}$ and $s_{c,t}$. Therefore the literature tends to distinguish between static and dynamic agglomeration effects. When agglomeration effects are static, $B_{c,t}$ is immediately affected by current values of local characteristics but not by earlier values. This means that a larger city size in a given year affects local productivity only that year, and that any future change in city size will instantaneously translates into a change in local productivity. By contrast, recent contributions simultaneously consider some possible long-lasting effects of local characteristics that are called dynamic effects. We focus here on static affects and introduce dynamic effects from Section 2.2 onwards.

Let us turn now to a first empirical specification encompassing static agglomeration effects where the logarithm of the composite productivity effect, $B_{c,t}$, is specified in reduced form as a function of the logarithm of local characteristics and some local unobserved effects. Average local skills, $s_{c,t}$, are specified as a log-linear function of local education and again some local unobserved terms. The sum of all unobserved components is supposed to be a random residual denoted $\eta_{c,t}$. Denoting $y_{c,t}$ the mesure of the local outcome, here the logarithm of local wage, we obtain from

equation (3) the specification:

$$y_{c,t} = Z_{c,t}\gamma + \eta_{c,t}, \tag{4}$$

where $Z_{c,t}$ includes local variables for both the composite productivity effect and local skills. If explanatory variables reduce to the logarithm of density, it is assumed that local skills variables capture only skill composition effects and that there is no correlation between the random component and density, then the ordinary least square (OLS) estimate of the elasticity of productivity with respect to density, γ, is a consistent measure of total net agglomeration economies. This elasticity is crucial from the policy perspective even if the channels of agglomeration and dispersion forces are not identified. For instance, a value for γ of 0.03 means that a city twice as large (knowing that a factor of 10 is often obtained for the inter-quartile of local density in many countries) has $2^{0.03} - 1 \approx 2.1\%$ greater productivity, because of either pure local externalities or market agglomeration effects that dominate dispersion effects of any kind.

As mentioned in the introduction, the usual goal of the empirical works is to identify causal impacts, i.e. what would be the effect on local outcomes of changing some of the local characteristics. Beyond other endogeneity concerns discussed below, a first issue with specification (4) is that density can be correlated with some of the local unobserved skill components entering the residual. For instance, proxies for local skills such as diplomas may not be enough to capture all the skills that affect productivity. If unobserved skills are randomly distributed across locations, the OLS estimate of γ is a consistent estimator of the magnitude of agglomeration economies. Alternatively, if unobserved skills are correlated with density, there is an endogeneity issue and the OLS estimate of γ is biased.

Unobserved skills can be taken into account with individual panel data. This requires to extend our setting to the case where workers are heterogeneous. We assume now that local efficient labour is given by the sum of all efficient units of labour provided by heterogeneous workers, i.e. $s_{c,t}L_{c,t} = \sum_{i\in\{c,t\}} s_{i,t}\ell_{i,t}$ where $\ell_{i,t}$ is the number of working hours provided by individual i and $s_{i,t}$ is individual efficiency at date t. The wage bill is now $\sum_{i\in\{c,t\}} \omega_{i,t}\ell_{i,t}$, where $w_{i,t}$ is the individual wage. Profit maximization leads to

$$w_{i,t} = B_{c,t}s_{i,t}. \tag{5}$$

Let $X_{i,t}$ be time-varying observed individual characteristics and u_i an individual fixed effect to be estimated. We make the additional assumption that individual efficiency can be written as the product of an individual-specific component $\exp(X_{i,t}\theta + u_i)$ and of a residual $\exp(\epsilon_{i,t})$ reflecting individual- and time-specific random effects. Here, u_i captures the effects of individual unobserved skills which are supposed to be constant over time. Taking the logarithm of (5) and using the same specification of agglomeration effects as for (4) gives:

$$y_{i,t} = u_i + X_{i,t}\theta + Z_{c(i,t),t}\gamma + \eta_{c(i,t),t} + \epsilon_{i,t}\,, \tag{6}$$

where $y_{i,t}$ is the individual local outcome, here the logarithm of individual wage at date t, and $c(i,t)$ is the labour market where individual i is located at date t. Note that we implicitly assume a homogeneous impact of local characteristics γ across all workers, areas and industries. Heterogeneous impacts are considered in Subsection 2.1.2. For now, we consider that individual fixed effects are here only to capture unobservable skills although we will discuss in Subsection 2.2 the fact that they can also capture learning effects that may depend on city size.

The use of individual data and the introduction of an individual fixed effect in specification (6) were first proposed by Glaeser and Maré (2001), and this should largely reduce biases due to the use of imperfect measures of skills. Most importantly, the individual fixed effect makes it possible to control for all the characteristics of the individual shaping their skills that do not change over time and the effect of which can be considered to be constant over time. They include education, which is often observable, but also many other characteristics that are more difficult to observe such as the education of parents and grand-parents, the number of children in the family, mobility during childhood, or personality traits. Since the individual fixed effects are allowed to be correlated with local variables such as density, one can more safely conclude that the effects of local characteristics do not capture some composition effects due to sorting on the individual characteristics.

The second advantage of individual data is that the local average of any observed individual characteristic can be introduced in the set of local variables simultaneously with the individual characteristic itself if it is time-varying or with the individual fixed effect. In particular, while the individual fixed effect controls for the individual level of education, one can consider in $Z_{c,t}$ the local share of any education level to assess whether highly skilled workers exert a human capital local externality on other workers.[2] The estimated effects of local variables such as density then correspond to agglomeration economies other than education externalities. As discussed above, such a distinction cannot be made when using aggregate data.

The sources of identification of local effects can be emphasised by considering specification (6) in first difference, which makes the unobserved individual effect disappear. For simplicity's sake, consider only two terms in the individual outcome specification such that $y_{i,t} = Z_{c(i,t),t}\gamma + u_i$ where $Z_{c,t}$ only includes density. For individuals staying in the same local market c at two consecutive dates, the first difference of outcome is given by $y_{i,t} - y_{i,t-1} = (Z_{c,t} - Z_{c,t-1})\gamma$, and time variation of density within the local market for instance participates to the identification of the density effect, γ. For individuals moving from market c to market c', we have: $y_{i,t} - y_{i,t-1} = \left(Z_{c',t} - Z_{c,t-1}\right)\gamma$, and both spatial and time variations of density contribute to identifying the density effect. If there is no mover, agglomeration economies are still identified, but from time variations for stayers only. This is because there is a single parameter to estimate and averaging the first-differenced outcome equation of stayers at the local-time level, one gets $Z \times (T-1)$ independent relationships where

[2]The interpretation based on externalities requires further caution. It is discussed in Section 3.3.

Z is the number of local markets.

Note that we assume for the moment that the specification is the same for stayers and movers, i.e. that the individual parameters θ, the effects of local characteristics γ, and the distributions of random components are identical. Should this assumption be questioned, one could choose to estimate (6) separately on the subsamples of stayers and movers since identification is granted for each sub-sample, and one could in turn use the separate estimates to test the assumption of homogeneity between the two groups.

Specification (6) can be estimated directly by OLS once written in first difference (or projected in the within-individual dimension) to remove the individual fixed effects, but the computation of standard errors is an issue. Indeed, the covariance matrix has a complex structure due to unobserved local effects and the mobility of workers across labour markets. For mobile individuals, the first difference of the specification includes two different unobserved local shocks, $\eta_{c',t}$ and $\eta_{c,t-1}$, and the locations of those shocks (c and c') vary across mobile individuals, even for those initially in the same local market because they may not have the same destination after they move. There is thus no way to sort individuals properly to get a simple covariance matrix structure and to cluster standard errors at each date by location. It is tempting to ignore unobserved local effects but this can lead to important biases of the estimated standard errors for effects of local variables as shown by Moulton (1990).

Alternatively, it is possible to use a two-step procedure that both solves this issue and has the advantage of corresponding to a more general framework. Consider the following system of two equations:

$$y_{i,t} = u_i + X_{i,t}\theta + \beta_{c(i,t),t} + \epsilon_{i,t}\,, \tag{7}$$

$$\beta_{c,t} = Z_{c,t}\gamma + \eta_{c,t}\,, \tag{8}$$

where $\beta_{c,t}$ is a local-time fixed effect that captures the role of any location-time variable whether it is observed or not. The introduction of such fixed effects capturing local unobserved components makes the assumption of independently distributed shocks more plausible. The specification is also more general since it takes into account possible correlations between local-time unobserved characteristics and individual characteristics. There are thus fewer possible sources of biases and this in turn should lead to a more consistent evaluation of the role of local characteristics.

Estimating this model is more demanding in terms of identification, and having movers between locations is now required. Assume for simplicity's sake that the first equation of the model is given by $y_{i,t} = \beta_{c(i,t),t} + u_i$. When one rewrites this specification in first difference for non-movers and movers, one gets $y_{i,t} - y_{i,t-1} = \beta_{c,t} - \beta_{c,t-1}$ and $y_{i,t} - y_{i,t-1} = \beta_{c',t} - \beta_{c,t-1}$, respectively. There is one parameter $\beta_{c,t}$ to be identified for each location at each date. If there is no mover, one wishes to average the specification at the local-time level for stayers as before but ends up with

$(Z-1) \times T$ independent relationships whereas there are $Z \times T$ parameters to estimate. In other words, one can identify the time variations of local effects for any location but not their differences between locations.

By contrast, when there are both stayers and movers, identification is granted as it can be shown in a way similar to a difference-in-differences approach. The difference of the wage time variation between a mover to c', denoted i', and a non-mover i initially in the same location c is given by $\big(y_{i',t} - y_{i',t-1} - (y_{i,t} - y_{i,t-1})\big) = \beta_{c',t} - \beta_{c,t}$. For any pair of locations, the difference in wage growth between movers and non-movers identifies the difference of local effects between the two locations. Moreover, the wage growth of stayers identifies the variation of local effects over time as before. All parameters $\beta_{c,t}$ are finally identified when local markets are well inter-connected through stayers and flows of movers, up to one that needs to be normalised to zero as differences do not allow the identification of levels. Inter-connection means that any pair of location-time couples, (c,t) and (c',t'), can be connected through a chain of pairs of location-time couples (j,τ) and $(j',\tau+1)$ such that there are migrants from j to j' between dates τ and $\tau+1$ if $j \neq j'$, or stayers in j between the two dates if $j = j'$.[3] In other words, assuming that there are some migrants between every pair of locations in the dataset, we have $Z^2 \times (T-1)$ independent relationships and only $Z \times T - 1$ parameters to estimate. Crucially, the assumption that the specification is identical for both movers and stayers is now required, otherwise identification is not possible. Alternatively, more structural approaches can help to some extent to solve the identification issue and we present them in Section 2.4.

Note finally that in practice specification (7) is estimated in a first step. Panel data estimation techniques such as within estimation are used because considering a dummy variable for each individual to take into account the fixed effect u_i would be too demanding for a computer. The estimates of $\beta_{c,t}$ are then plugged into equation (8). The resulting specification is estimated in a second stage using linear methods, including one observation for the location-time fixed effect normalised to zero. The sampling error on the dependent variable, which is estimated in first stage, must be taken into account in the computation of standard errors, and it is possible to use Feasible General Least Squares (FGLS) (see Combes et al., 2008a, for the implementation details). A more extensive discussion on the estimation strategy addressing endogeneity issues is presented in Section 4, but we first augment the model to consider the role of more sophisticated agglomeration mechanisms.

[3] If local markets are not all inter-connected, groups of fully inter-connected location-time couples must be defined ex-ante such that location-time fixed effects are all identified within each group up to one being normalised to zero. For more details, the reader may refer to the literature on the simultaneous identification of worker and firm fixed effects in wage equations initiated by Abowd et al. (1999).

2.1.2 Heterogeneous impact of local effects

The profit maximisation we conduct above to ground our specification emphasises that agglomeration effects may relate to pure externalities, or to good or input price effects. Obviously, the magnitude of these channels may differ across industries. For instance, the impact of density may be greater in high-tech industries due to greater technological externalities, and good or input price effects depend on the level of trade costs within each industry. The consideration of agglomeration mechanisms that are heterogeneous across industries simply requires extending the specification such that:

$$y_{i,t} = u_i + X_{i,t}\theta + Z_{c(i,t),t}\gamma_{s(i,t)} + \eta_{c(i,t),s(i,t),t} + \epsilon_{i,t} \tag{9}$$

where $s(i,t)$ is the industry where individual i works at time t, γ_s is the effect of local characteristics in industry s, and $\eta_{c,s,t}$ is a location-industry-time shock. This specification can be estimated in several ways. The most straightforward one consists of splitting the sample by industry and implementing the approach proposed in Section 2.1.1 for each industry separately. Nevertheless this means that the coefficients of individual explanatory variables as well as individual fixed effects are not constrained to be the same across industries, which may or may not be relevant from a theoretical point of view. This also entails a loss of precision for the estimators. An alternative approach consists of considering among explanatory variables some interactions between density, or any other local characteristic, and industry dummies, and estimating the specification in the within-individual dimension as before to recover their coefficients which are the parameters γ_s.

Again estimated standard errors may be biased due to heteroskedasticity arising from location-industry-time random effects, $\eta_{c,s,t}$. To deal with this issue, it is possible to consider a two-step approach which makes use of location-industry-time fixed effects, $\beta_{c,s,t}$, in the following system of equations:

$$y_{i,t} = u_i + X_{i,t}\theta + \beta_{c(i,t),s(i,t),t} + \epsilon_{i,t} \tag{10}$$

$$\beta_{c,s,t} = Z_{c,t}\gamma_s + \eta_{c,s,t} \tag{11}$$

Location-industry-time fixed effects are estimated with OLS once equation (10) has been projected in the within-individual dimension, as previously when estimating location-time fixed effects. They are identified up to one effect normalised to zero provided that all locations and industries are well inter-connected by workers mobile across locations and industries.[4] Their estimators are plugged into equation (11) which is estimated in a second step.

Importantly, introducing the industry dimension increases the number of local characteristics that can have an agglomeration effect. It has become common practice to distinguish between

[4] As before, groups of fixed effects should be defined ex-ante if not all locations and industries are properly inter-connected. Of course, the larger the number of industries, the more likely location-industry-time fixed effects are not all identified.

urbanisation economies and localisation economies. Whereas urbanisation economies correspond to externalities arising from characteristics of the location such as density, localisation economies correspond to externalities arising from characteristics of the industry within the location. The determinants of agglomeration economies considered in the literature thus depend only on location for urbanisation economies and on both location and industry for localisation economies. The local determinant of localisation economies most often considered is specialisation, which is defined as the share of the industry in local employment. While the use of density makes it possible to assess whether productivity increases with the overall size of the local economy, the use of specialisation allows the assessment of whether it increases with the local size of the industry in which the firm or worker operates. The pure externalities and market externalities distinguished above can operate at the whole location scale or at the industry-location level. In line with these arguments, one may rather want to estimate in the second stage the following specification:

$$\beta_{c,s,t} = Z_{c,t}\gamma_s + W_{c,s,t}\delta_s + \eta_{c,s,t} \tag{12}$$

where $W_{c,s,t}$ are determinants of localisation economies including specialisation and $Z_{c,t}$ are the determinants of urbanisation economies. All the local characteristics considered in the literature are detailed in Section 3.

One estimation issue is that the number of fixed effects to estimate in the first stage increases rapidly with the number of locations and we are not aware of any attempt to estimate the proposed specification. As an alternative, one can mix strategies as proposed by Combes et al. (2008a) and estimate:

$$y_{i,t} = u_i + X_{i,t}\theta + \beta_{c(i,t),t} + W_{c(i,t),s(i,t),t}\delta_{s(i,t)} + \epsilon_{i,t} \tag{13}$$

$$\beta_{c,t} = Z_{c,t}\gamma + \eta_{c,t} \tag{14}$$

This model is less general than (10) and (12) since unobserved location-industry-time effects are not controlled for in the first step and determinants of urbanisation economies are assumed to have a homogeneous impact across industries in the second step (as γ does not depend on industry). Still, heterogeneous effects of determinants of localisation economies are identified in the first stage on top of controlling for unobserved location-time effects.

It is also easy to argue from theory that agglomeration effects are heterogeneous across different types of workers. Some evidence suggests for instance that more productive workers are also the ones more able to reap the benefits from agglomeration (see for instance Glaeser and Maré, 2001; Combes et al., 2012c; De La Roca and Puga, 2012). A specification similar to (9) can be used to study for instance the heterogeneous effect of density across diplomas. One would simply consider diploma-specific coefficients for density instead of industry-specific ones. However, diplomas usually do not change over time. When using a two-step procedure, this implies that

one local-diploma-time fixed effect must be normalised to zero for each diploma. The alternative strategy of estimating the two-step procedure on each diploma separately is not much less precise than it was for industries since all the observations of any given individual are in the same diploma sub-sample, and there is thus a unique individual fixed effect for each worker to be estimated.

However diplomas may not be enough to fully capture individual skill heterogeneity. One may wish to consider that the effect of density is specific to each individual as in the following specification:

$$y_{i,t} = u_i + X_{i,t}\theta + Z_{c(i,t),t}\gamma_i + \eta_{c(i,t),t} + \epsilon_{i,t} \tag{15}$$

where γ_i is an individual fixed effect. Parameters can be estimated using an iterative procedure.[5] For a given value of θ, one can regress $y_{i,t} - X_{i,t}\theta$ on $Z_{c(i,t),t}$ for each individual. This gives some estimates for γ_i and u_i. Then, θ is estimated by regressing $y_{i,t} - Z_{c(i,t),t}\gamma_i - u_i$ on $X_{i,t}$. The procedure is repeated using the parameter values from previous iteration until convergence.

One can further extend the model and consider that location in general, and not density alone, has a heterogeneous effect on the local outcome. One considers in this case an interaction term between a local fixed effect and an individual fixed effect. This amounts to saying that it is not the effect of density but rather the combined effect of all local characteristics, whether they are observed or not, which is heterogeneous across individuals. The first step of the two-step procedure in this case becomes:

$$y_{i,t} = u_i + X_{i,t}\theta + \beta_{c(i,t),t} + \delta_{c(i,t),t}v_i + \epsilon_{i,t} \tag{16}$$

with the identification restriction that $\sum_i v_i = 0$ and one of the local terms $\delta_{c,t}$ is normalised to zero. As before, the specification can be estimated with an iterative procedure. The estimators of parameters $\delta_{c,t}$ are regressed in second step on local variables to assess the extent to which agglomeration economies influence the local return of unobserved individual characteristics. An additional extension to make the specification even more complete would consist of having the coefficients of individual characteristics depend on the individual. Note that as there are many individual-specific effects entering the model in a non-additive way, the time span should be large for the estimations to make sense and there is no guarantee that a large number of periods is enough for the parameters to be properly estimated. In any case, most specifications in this subsection are material for future research.

2.2 Dynamic impact of agglomeration economies

So far, we have considered that agglomeration economies have an instantaneous effect on productivity and then no further impact in the following periods. In fact, agglomeration economies can be dynamic and have a permanent impact such as when technological spillovers increase local productivity growth or when individuals learn more or faster in larger cities as suggested by Lucas

[5] This procedure is inspired from Bai (2009) who proposes such a procedure to estimate factor models.

(1988). One can even argue that when an individual moves from a large city to a smaller one, they can transfer part of their productivity gains from agglomeration to the new location and be more productive than other individuals who have not worked in a large city. In that case, dynamic effects operate through the impact of local characteristics on the growth of $A_{c,t}$ and $s_{i,t}$ which are involved in equation (5). One can also consider dynamic effects operating through $p_{c,t}$ and $r_{c,t}$. For instance, agglomeration can facilitate the diffusion of information about the quality of goods and inputs, and this in turn can have an impact on price variations across periods (for instance when prices are chosen by producers under imperfect competition). Therefore, even if dynamic effects relate more plausibly to technological spillovers and learning effects, market agglomeration economies can also present dynamic features. As a result, identification issues are like those for static agglomeration economies and one usually estimates only the overall impact of dynamic externalities and not the exact channel through which they operate. Note that the literature that first tried to identify agglomeration effects on local industrial employment, which dates back to Glaeser et al. (1992) and Henderson et al. (1995), adopts this dynamic perspective from the very beginning. We present this literature in Section 6.1.

We explain in this section how the previous productivity specifications can be extended to encompass dynamic effects. The distinction between static and dynamic effects was pioneered by Glaeser and Maré (2001) and we elaborate the discussion below from their ideas and those developed by De La Roca and Puga (2012) which is currently one of the most complete studies on the topic. For a model with static local effects only (ignoring the role of time-varying individual and industry characteristics), written as $y_{i,t} = u_i + \beta_{c(i,t),t} + \epsilon_{i,t}$, the individual productivity growth rate is simply related to the time difference of static effects:

$$y_{i,t} - y_{i,t-1} = \beta_{c(i,t),t} - \beta_{c(i,t-1),t-1} + \varepsilon_{i,t}, \tag{17}$$

where $\varepsilon_{i,t}$ is an error term.[6] Dynamic local effects in their simplest form are introduced by assuming for $t \geq 1$:

$$y_{i,t} - y_{i,t-1} = \beta_{c(i,t),t} - \beta_{c(i,t-1),t-1} + \mu_{c(i,t-1),t-1} + \varepsilon_{i,t}. \tag{18}$$

where $\mu_{c,t-1}$ is a fixed effect for city c at date $t-1$ which corresponds to city c impact on productivity growth between $t-1$ and t, and thus captures dynamic local effects. Interestingly, this implies:

$$y_{i,t} = y_{i,0} + \beta_{c(i,t),t} + \sum_{k=1}^{t} \mu_{c(i,t-k),t-k} + \zeta_{i,t}, \tag{19}$$

where $\zeta_{i,t}$ is an error term. This equation includes the past values of local effects and shows that dynamic effects, even when they affect only the annual growth rate of a local outcome, do have a

[6] In this article, we consider that $\varepsilon_{i,t}$ is a generic notation for the residual and use it extensively in different contexts.

permanent impact on its level. Nevertheless, we have made some major assumptions to reach this specification. We now detail them and discuss how to relax them.

A first implicit assumption is that dynamic effects are perfectly transferable over time. For instance, knowledge does not depreciate even after a few years. To consider depreciation, one could introduce in (18) some negative effects of past city terms $\mu_{c(i,t-1),t-k}$, $k > 1$ with coefficients lower than one in absolute value, and this would lead to an auto-regressive specification such that terms $\mu_{c(i,t-1),t-k}$ have an effect attenuated with a time lag when the model is rewritten in level.

Importantly, specification (19) makes more sense for individuals who stay in the same location than for movers. Dynamic local effects might also depend on where individuals locate at period t, and therefore on the destination location for movers. Individuals in a large city probably do not benefit from the same productivity gains from learning effects whether they move to an even larger city or to a smaller city (or if they stay where they are). In other words, dynamic gains are not necessarily fully transferable between locations, and the degree of transferability can depend on the characteristics of locations. Therefore, it might be more relevant to assume that dynamic effects depend on both the origin and destination locations and rewrite the specification of local outcome as:

$$y_{i,t} = y_{i,0} + \beta_{c(i,t),t} + \sum_{k=1}^{t} \mu_{c(i,t-k),c(i,t),t-k} + \zeta_{i,t}. \tag{20}$$

where $\mu_{j,c,\tau}$ is a fixed effect for being in city j at date $\tau < t$ and in city c at date t. The problem is that the number of parameters to be estimated for dynamic effects becomes very large (the square of the number of locations times the number of years in the panel). Moreover, restrictions on parameters must be imposed for the model to be identified. This can be seen for instance when writing the model in first difference for workers staying in the same location between dates $t-1$ and t, for which $c(i,t-1) = c(i,t)$:

$$y_{i,t} - y_{i,t-1} = \beta_{c(i,t),t} - \beta_{c(i,t-1),t-1} + \mu_{c(i,t-1),c(i,t),t-1} + \varepsilon_{i,t}. \tag{21}$$

The evolution of the static agglomeration effect cannot be distinguished from the dynamic effect (and this is also true when considering movers instead of stayers). When one observes the productivity variation of stayers, one does not know whether it occurs because static local effects changed or because some dynamic local effects take place.

De La Roca and Puga (2012) make some assumptions that grant the identification of the model and significantly reduce the number of parameters to be estimated. They assume that static and dynamic effects do not change over time, i.e. $\beta_{c,t} = \beta_c$ and $\mu_{j,c,t-k} = \mu_{j,c}$. Under these assumptions, $\mu_{c,c}$ captures both the dynamic effect and the evolution of static effects. This can be seen from equation (21) where the evolution of static effects would be now fixed to zero. This should be kept in mind when assessing the respective importance of static and dynamic effects, as this cannot be done from the relative explanatory power of β_c and $\mu_{j,c}$. Under these assumptions,

it is also possible to rewrite the specification in a more compact form introducing the number of years the individuals have spent in each location:

$$y_{i,t} = u_i + X_{i,t}\theta + \beta_{c(i,t)} + \sum_j \mu_{j,c(i,t)} e_{i,j,t} + \epsilon_{i,t}, \tag{22}$$

where $e_{i,j,t}$ is the experience acquired by individual i until period t in city j (the number of years they spent there until date t), and $\mu_{j,c}$ captures the value of one year of this experience when the worker is located in city c. One can test whether the $\mu_{j,c}$ are statistically different from each other when c varies for given j, i.e. whether location-specific experience can be transferred or not to the same extent to any location, as was assumed in (19). One can also quantify the respective importance of the effects β_c and $\mu_{c,c}$ keeping in mind that it does not correspond to the respective importance of static and dynamic effects. Earlier attempts to evaluate dynamic effects on wages by Glaeser and Maré (2001), Wheeler (2006) and Yankow (2006) correspond to constrained and simplified versions of this specification, typically distinguishing only the impact on wage growth of moving or not moving to larger cities.

It is then possible in a second stage to evaluate the extent to which dynamic effects depend on the characteristics of the local economy, and assess whether transferability relates to the density of the destination location. One can consider the specification:

$$\mu_{j,c} = Z_{j,\bullet}\left(\psi + Z_{c,\bullet}\upsilon\right) + \zeta_{j,c} \tag{23}$$

where $Z_{j,\bullet}$ is the average over all periods of a vector of location-j characteristics including density. In this specification, the effect of density in the location where learning took place is a linear function of variables entering $Z_{c,\bullet}$ such as density. Clearly, all these dynamic specifications can be extended to encompass some heterogeneity across industries in the parameters of local variables, and possibly some localisation effects.

An alternative approach that takes into account time variations in static and dynamic effects may consist of estimating density effects in one stage only, first specifying:

$$\beta_{c,t} = Z_{c,t}\gamma + \eta_{c,t} \tag{24}$$

$$\mu_{j,c,t} = Z_{j,t}\left(\psi + Z_{c,t}\upsilon\right) + \zeta_{j,c,t}, \tag{25}$$

and then plugging these expressions into equation (20). This gives a specification where coefficients associated with the different density terms can be estimated directly with linear panel methods. A limitation of this approach is again that it is difficult to compute standard errors taking into account unobserved local shocks because workers' moves make the structure of the covariance matrix of error terms intricate when the model is rewritten in first difference or in the within dimension. On the other hand, the separate explanatory power of static and dynamic agglomeration effects is

better assessed.

Finally, it is possible to generalise the framework to the case where both static and dynamic effects are heterogeneous across individuals. Specification (20) becomes:

$$y_{i,t} = u_i + X_{i,t}\,\theta + \beta_{c(i,t),t} + \delta_{c(i,t),t}\,v_i + \sum_{k=1}^{t} \left(\mu_{c(i,t-k),c(i,t),t-k} + \lambda_{c(i,t-k),c(i,t),t-k}\,r_i\right) + \epsilon_{i,t}, \qquad (26)$$

where v_i and r_i are individual fixed effects verifying the identification assumption $\sum_i v_i = \sum_i r_i = 0$. Parameters can be estimated imposing additional identification restrictions such as the fact that static and dynamic effects do not depend on time, and using an iterative procedure as in previous subsections. Note that such a specification has not been estimated yet. One of the closest attempts is De La Roca and Puga (2012) who restrict the spatial dimension to three classes of city sizes only (which prevents the second stage estimation and only allows them to compare the experience effect over the three classes). Importantly, they also make the further (strong) assumption that the impact of individual heterogeneity is identical for both static and dynamic effects, i.e. $v_i = r_i$. D'Costa and Overman (2014) is another attempt elaborating on De La Roca and Puga (2012). They estimate the specification in first differences while allowing for $v_i \neq r_i$, but they exclude movers to avoid having to deal with between-city dynamic effects.

2.3 Extending the model to local worker-firm matching effects

Marshall (1890) was among the first to emphasise that agglomeration can increase productivity by improving both the quantity and quality of matches between workers and firms on local labour markets (see Duranton and Puga, 2004, for a survey of this type of mechanism). The better average quality of matches in larger cities can be considered as a static effect captured by the local fixed effects $\beta_{c,t}$ estimated in previous subsections. The matching process in cities can also yield more frequent job changes, which can boost productivity growth. This dynamic matching externality can be incorporated into our framework considering that at each period t, a worker located in c receives a job offer with probability ϕ_c to which is associated a wage $\tilde{y}_{i,t}$. One assumes that workers change jobs within the local market at no cost and they accept a job offer if the associated wage is higher than the one they would get if they stayed with the same employer. To ease exposition, we suppose that migrants do not receive any job offer at their origin location but receive one at the destination location once they have migrated. The probability of receiving such an offer is supposed to be the same as that of stayers in this market. We also assume for the moment that there is no other dynamic effect than through job change. For workers receiving an offer, the wage at time t is $y_{i,t} + \Delta_{i,t}$ where $y_{i,t}$ is given by equation (7) and $\Delta_{i,t} = \max(0, \tilde{y}_{i,t} - y_{i,t})$. The

individual outcome is then given by:

$$y_{i,t} = u_i + X_{i,t}\theta + \beta_{c,t} + \sum_{\tau=1}^{t-1} 1_{\{O(i,\tau)=1\}}\Delta_{i,\tau} + \epsilon_{i,t} \tag{27}$$

where $O(i,\tau)$ is a dummy variable taking the value one if individual i has received a job offer between dates $\tau - 1$ and τ, and zero otherwise.

For workers keeping the same job in location c between the two dates, there is no dynamic matching gain and wage growth verifies:

$$y_{i,t} - y_{i,t-1} = (X_{i,t} - X_{i,t-1})\,\theta + \beta_{c,t} - \beta_{c,t-1} + \varepsilon_{i,t}, \tag{28}$$

where $\varepsilon_{i,t} = \epsilon_{i,t} - \epsilon_{i,t-1}$.

For workers changing jobs within location c, improved matching induces a wage premium $\Delta_{i,t}$ and wage growth can be written as:

$$y_{i,t} - y_{i,t-1} = (X_{i,t} - X_{i,t-1})\,\theta + \widetilde{\beta}_{c,t} - \beta_{c,t-1} + \nu_{i,t}, \tag{29}$$

where $\widetilde{\beta}_{c,t} = \beta_{c,t} + E\left(\Delta_{i,t} | \, i \in (c,t-1), i \in (c,t)\right)$ is the sum of the local fixed effect for stayers keeping their jobs and the expected productivity gain when changing job, and the new residual is $\nu_{i,t} = \varepsilon_{i,t} + \Delta_{i,t} - E\left(\Delta_{i,t} | \, i \in (c,t-1), i \in (c,t)\right)$.

For workers changing job between two locations c and c', wage growth can be expressed as:

$$y_{i,t} - y_{i,t-1} = (X_{i,t} - X_{i,t-1})\,\theta + \beta^{c}_{c',t} - \beta_{c,t-1} + \nu_{i,t}, \tag{30}$$

where $\beta^{c}_{c',t} = \beta_{c,t} + E\left(\Delta_{i,t} | \, i \in (c,t-1), i \in (c',t)\right)$ is the sum of the local fixed effect for stayers keeping their jobs in the destination location and the expected productivity gain when changing jobs from city c to city c'.[7] This gain may depend on both cities as it could be related for instance to the distance between them or their industrial structure.

The difference in local effects from separate wage growth regressions for stayers changing jobs and stayers keeping the same job provides an estimate of the matching effect since $\widetilde{\beta}_{c,t} - \beta_{c,t} = E\left(\Delta_{i,t} | \, i \in (c,t)\right)$. If changing jobs increases productivity through improved matching, this difference should be positive for any location c. If agglomeration magnifies such dynamic matching effects, the probability of changing jobs should increase with density, and the difference $\widetilde{\beta}_{c,t} - \beta_{c,t}$ should be larger in denser areas. More generally, to assess which local characteristics are determi-

[7] In fact, workers may move and take a wage cut if they expect future wage gains. This kind of inter-temporal behaviour cannot be taken into account in a static model as here but it can in the dynamic framework developed in the next subsection.

nants of dynamic matching effects, one can run the second-step regression:

$$\widetilde{\beta}_{c,t} - \beta_{c,t} = Z_{c,t}\Phi + \eta_{c,t}\,. \tag{31}$$

where $Z_{c,t}$ is a vector of local characteristics. Such a model has not been estimated yet but Wheeler (2006) is one of the closest attempts to do so. Due to the small size of its data set, the author of this article cannot identify the role of local time fixed-effects but his strategy on the panel of workers changing job is equivalent to directly plugging (31), with local market size as the single local characteristic, into the difference between (28) and (29) to assess by how much the matching effect increases with local market size.

Exploiting wage growth for workers changing both job and city is more intricate, and an important assumption which needs to be made (and was implicitly made in previous sections) is that the location choice is exogenous. In order to get consistent estimates of local effects when movers are used as a source of identification, the location choice should not depend on individual-location shocks on wages conditional on all the explanatory variables and parameters in the model[8]. This assumption is disputable since workers often migrate because they receive a good job offer in another local labour market, or because they had a bad original match with their firm. By the same token, we can argue that job changes are endogenous for both movers and non-movers, and this affects the estimates of local effects obtained for specifications in this sub-section. As this concern is certainly important, it may be wise to use another kind of approach that explicitly takes into account the endogeneity of location and job choices. This can be done with a dynamic model at the cost of imposing more structure on the specification that is estimated. We now turn to this kind of structural approach, building on the same underlying background.

2.4 Endogenous inter-temporal location choices

So far, we have considered static and dynamic agglomeration effects within a static framework where workers' location choices are strictly exogenous: Workers do not take into account wage shocks due to localised job opportunities in their migration or job change decisions. When workers do consider alternative job opportunities when making their decisions, it is also likely that they are forward-looking and take into account all future possible outcomes in alternative locations. As shown by Baum-Snow and Pavan (2012), it is possible to introduce static and dynamic agglomeration effects in a dynamic model of location choices that takes these features into account.[9] Nevertheless, identification is achieved thanks to the structure of the model and it is sometimes difficult to assess which conclusions would remain under alternative assumptions. For simplicity's

[8]This assumption is discussed at greater length from an econometric point of view in Section 4.2.

[9]Gould (2007) also proposes a dynamic model where school attendance too is endogenous. See also Beaudry et al. (2014) for a dynamic model with search frictions and wage bargaining with static agglomeration effects but no dynamic agglomeration effects.

sake, we present the main mechanisms of the model for employed workers and consider that there is no unemployment and no consumption amenities, these assumptions being relaxed in Baum-Snow and Pavan (2012). Unemployment can easily be added considering that there is an additional state for workers and there are exogenous mechanisms (such as job destruction and job offers) leading to transitions between states. Consumption amenities can be considered by including location-specific utility components that do not affect local wages.

Individual unobserved heterogeneity is modeled as draws in a discrete distribution (instead of individual fixed effects). There are H types of workers indexed by $h = 1, ..., K$. Worker i getting a job in location c draws a match $\varsigma_{i,c}$ in a distribution which is specific to the location. For a given job, the match is drawn once and for all and does not vary over time. The wage of worker i of type $h(i)$ located in market c and occupying a job with match $\varsigma_{i,c}$ is a variant of equation (22) given by:

$$y_{i,c,t}(\varsigma_{i,c}) = X_{i,t}\theta + \beta_{h(i),c,t} + \sum_j \mu_{h(i),j,c}\, e_{i,j,t} + \varsigma_{i,c} + \epsilon_{i,c,t} \tag{32}$$

where $\beta_{h,c,t}$ is a static location effect depending on the worker type, $\mu_{h,j,c}$ is a location-specific experience effect depending on the worker type, and $\epsilon_{i,c,t}$ is white noise. Note that whereas wage depends on the draw of the white noise, we do not index wage by it to keep notations simple. A crucial difference from the specifications in previous sections is that we now have a specification for the potential outcome in any location c at each date. Therefore, the wage is now indexed by c and we write $y_{i,c,t}$ for any potential wage instead of $y_{i,t}$ previously for the realised one.

The inter-temporal utility and location choice are determined in the following way. Consider worker i of type $h(i)$ located in city c at period t. The worker earns a wage $y_{i,c,t}$ and, at the end of the period, has the possibility to move to another job within the same location or to a different location. Migration to another location can be achieved only if the worker gets a job offer in that location (as we have ruled out unemployment for simplicity). The probability of receiving a job within location c for a worker of type h is denoted $\phi_{h,c}$, and the probability of receiving a job in location $j \neq c$ is denoted $\phi_{h,c,j}$. There is a cost C when changing job within the local market. If the worker moves between city c and city j, they have to pay a moving cost $M_{c,j}$. Let us denote $V_{i,c,t}(\varsigma_{i,c})$ the intertemporal utility of an individual located in city c at time t, and occupying a job with match $\varsigma_{i,c}$. This inter-temporal utility verifies the recursive formula:

$$\begin{aligned} V_{i,c,t}(\varsigma_{i,c}) &= y_{i,c,t}(\varsigma_{i,c}) + \phi_{h(i),c}\, E_{\varsigma_c} \max\left[V_{i,c,t+1}(\varsigma_{i,c}), V_{i,c,t+1}(\varsigma_c) - C\right] \\ &+ \sum_{j\neq c} \phi_{h(i),c,j}\, E_{\varsigma_j} \max\left[V_{i,c,t+1}(\varsigma_{i,j}), V_{i,j,t+1}(\varsigma_j) - M_{c,j}\right] \end{aligned} \tag{33}$$

where expectations are computed over the distributions of all future random terms including the matches ς_c when one changes jobs within location and ς_j when one changes jobs by moving to j (but not the realised match $\varsigma_{i,c}$ for current job). The first term corresponds to the wage earned

at current location. The second term is the expected outcome associated with a possible job offer within current location. It depends on the probability of receiving a job offer and on the expected future inter-temporal utility, which is the one related to the new job if it is worth accepting the offer, or the one related to the current job otherwise. The third term is the expected outcome associated with a possible job offer in other locations. It depends on the probability of receiving a job offer in every location and on the expected future inter-temporal utility related to the location if it is worth moving there, or to the current location otherwise.

The model can be estimated by maximum likelihood after writing the contributions to likelihood of individuals that correspond to their history of events (whether they change jobs, whether they change location, and their wages at each period). The model is parametrised by making some assumptions on the distributions of random and matching components, supposing they follow normal distributions with mean zero and variance to be estimated. Unobserved heterogeneity is modelled through mass points with individuals having some probabilities of being of every type which enter the set of parameters to be estimated. The computation of contributions to likelihood involves the integration over the distribution of unobserved components in line with Heckman and Singer (1984).

Once estimates of the parameters $\beta_{h,c,t}$, $\mu_{h,j,c}$, $\phi_{h,c}$ and $\phi_{h,c,j}$ are recovered, a variance analysis can be performed to assess the respective importance of static and dynamic local effects, as well as matching effects. Estimated parameters can also be regressed on density (or any other local variable), to evaluate how they vary with changes in the characteristics of locations. In practice, however, the numbers of locations and related parameters are usually too large for the model to be empirically tractable. An alternative is to aggregate locations by quartile of density and consider that each group is a single location in the model. Once parameters are estimated, it is possible to assess whether they take larger values for groups of denser locations.

Overall, structural approaches modelling jointly location choices and wages are an interesting tool for taking into account the endogeneity of workers' mobility when assessing the impact of local determinants of agglomeration economies, whereas this has never been properly done with linear panel models. Nevertheless, it comes at the cost of making strong assumptions about the structure of the model including parametric assumptions about random terms. More details on structural approaches in urban economics are provided by Holmes and Sieg (2015).

3 Local determinants of agglomeration effects

We have already argued that the literature usually estimates the total net impact of local characteristics related to agglomeration economies rather than the magnitude of agglomeration channels (although there are some tentative exceptions that are presented in Section 7). The previous section alludes to some of these local characteristics, in particular employment density. This section

details the definitions of all the characteristics that have been considered in the literature and explains to what extent they play a role in agglomeration economies. The outcome on which the impacts of local determinants of agglomeration economies are estimated often refers to a particular industry, either because data aggregated by location-industry are used or because one considers individual outcomes of firms or workers in a given industry. Considering this, two types of local characteristics may be included in the specification, those that are not specific to the industry and shape urbanisation economies, and those that are specific to the industry and shape localisation economies. We show successively how the size of the local market, the industrial structure of the local economy, and the composition of the local labour force can affect agglomeration economies and in turn local outcomes. We will see that in each case there can be both urbanisation and localisation economies.

3.1 Density, size and spatial extent of agglomeration effects

Equation (3) shows which pure and market agglomeration mechanisms involve the size of the local economy. Depending on the mechanism, employment, population, or production can be the most relevant variable to measure local economy size. However, the correlation between these three variables is often too great to allow the identification of their respective effects separately, and one has to restrict analysis to one of them. Results are in general very similar whichever variable is used. Employment is usually preferred to population, first because it better reflects the magnitude of local economic activity, and second because certain other local variables (described below) can be constructed from employment only. Production presents the disadvantage of being more subject to endogeneity issues than employment (see Section 4).

One usually considers models where both productivity and size are measured in logarithm because this eases interpretations, the estimated parameter being an elasticity. This also reduces the possibility of extreme values for the random component of the model and makes its distribution closer to the one of a normal law, which is usually used in significance tests.

Ciccone and Hall (1996) argue that the size of the local economy should be measured by the number of individuals per unit of land, i.e. density. Indeed, there is usually a large heterogeneity in the spatial extent of the geographic units that are used, as these units are often based on administrative boundaries. This can also create arbitrary border effects, an issue related to what the literature calls the Modifiable Areal Unit Problem (MAUP), i.e. the fact that some conclusions reached by empirical works could depend on the spatial classification used in their analyses, in particular the size and shape of the spatial units. Using density should reduce issues about mis-measurement of the size of the local economy, which is in line with Briant et al. (2010) who show that using more consistent empirical strategies largely reduces MAUP concerns.

Importantly, from the theory point of view, depending on the micro-foundations of pure et and market local externalities entering (3), either local density or the level of local employment

can affect the magnitude of the effects at stake. Therefore, there is no reason to restrict the specification to one variable or the other. Typically, if agglomeration gains outweigh agglomeration costs, one expects in general both density and size of the location to have a positive impact on local productivity. When considering variables in logarithm, it is possible and convenient to capture the two effects using density and land area simultaneously (while leaving employment aside). The impact of density, holding land area constant, reflects the gains from increasing either the number of people in the city or the density, while the impact of land area, holding density constant, reflects the gains from increasing the spatial extent of the city (i.e. from increasing both land area and employment proportionally). In a logarithmic specification, any combination of employment and land area identifies the same fundamental parameters but one has to be careful with the interpretation of coefficients, since we have:

$$\beta \ln den_{c,t} + \mu \ln area_{c,t} = \beta \ln emp_{c,t} + \varrho \ln area_{c,t} \text{ with } \varrho = \mu - \beta. \tag{34}$$

where $emp_{c,t}$ is total employment in location c at date t, $area_{c,t}$ is land area, and $den_{c,t} = \frac{emp_{c,t}}{area_{c,t}}$ is density. This equation shows that, whereas the effect of total employment for a given land area and the effect of density for a given land area correspond to the same parameter β, the effect of land area for a given total employment ϱ is equal to the difference between the effect of land area for a given density μ and the effect of density β. In fact, ϱ can be negative even when agglomeration gains result from both density and spatial extent. It would be wrong to conclude that there are agglomeration costs from a negative estimated value, or no agglomeration gains from spatial extent from a non-significant estimated coefficient. When using density and land area, agglomeration gains exist when any of the estimated coefficients is significantly positive.

Firms trade with distant markets, and communication exchanges occur between agents located sometimes pretty far apart. A number of studies have attempted to evaluate the spatial extent of local spillovers beyond the strict limits of the local unit. These spillovers can occur for any of the urbanisation and localisation effects considered in this section but most contributions in the literature consider them for local size only. Spatial econometric approaches usually consider spillovers for all the local determinants but at the cost of assuming for all of them an identical influence of distance on spillovers, and making it more difficult to deal with endogeneity issues (see Subsection 4.5.4). A flexible specification where density is considered at various distances from the worker's or firm's location may be envisaged. Typically, one can introduce in the specification many additional variables for density measured at 20, 50, 100, 150, 200 miles, etc., from the location. However, there is sometimes not enough variation in the data to identify so many effects of density. Therefore, some authors follow Harris (1954) and put more constraints on the impact of trade and communication costs by assuming that they increase with the inverse of distance,

which typically leads to Harris's following market potential variable:

$$MP_{c,t} = \sum_{\ell \neq c} \frac{den_{\ell,t}}{d_{c,\ell}}, \tag{35}$$

where $d_{c,\ell}$ is the distance between location c and location ℓ.

A number of variants for computing market potential exist since one can consider either population, employment or production, in level or in density form, as measures of market size. Several market potential variables can be considered simultaneously (for instance, one for density and one for land area). One can also refine the way trade and communication costs are assessed by using, instead of as-the-crow-flies distances, real distances by road or real measures of trade and communication costs. Nevertheless, all the corresponding market potential variables are usually highly correlated, as illustrated by Combes and Lafourcade (2005), and the effect of only one of them can actually be identified. If density is used as the measure of the local economy size, computing market potential using densities is more consistent. Importantly, the own location is excluded from formula (35) of Harris market potential to obtain an "external" market potential whose impact can usually be identified separately from the effect of the own location size. In any case, and as for own density, one cannot say whether the impact of market potential is a market-based effect or a pure externality, and more generally which mechanism is at play.

Fujita et al. (1999) emphasise that in economic geography models based on Dixit-Stiglitz monopolistic competition, local nominal wages are an increasing function of a specific variable, called the "structural market access", that closely relates to Harris market potential. Intuitively Dixit-Stiglitz models suggest that Harris's specification needs to be augmented with local price effects to take into account the role of imperfect competition that makes the price of the manufacturing good differ across locations due to its differentiation affecting both its supply and demand. In other words, there is now an impact of locations further away through $p_{c,t}$ in (3), which is captured by the structural market access variable. Note that the structural market access variable aggregates the effects of sizes of both own and distant locations, and its computation thus requires a consistent measure of trade costs not only between locations, but also within locations. This is a concern by itself as internal trade costs are usually not available in data sets and no fully satisfactory solution has been proposed yet to evaluate them. The most frequent strategy for coping with the issue, which is ad-hoc, consists of assuming that, within a location, trade costs are proportional to the square root of land area.

Interestingly, Redding and Venables (2004) show that in a model where varieties are used as intermediate inputs, another variable very similar to the market access, which may be called the "structural supply access", determines the price of inputs, $r_{c,t}$, in (3). The greater the supply access, the lower input prices and the higher nominal wages. Due to the strong link to the theory of the structural market access and the supply access, which makes them dependent on the elasticity

of substitution between varieties, for instance, no empirical counterpart can be directly constructed. Hanson (2005) was the first to suggest using also theory to relate market access to observables, and in particular local housing stocks. Redding and Venables (2004) take another route where both market and supply accesses are estimated through a first-step trade gravity equation, and their predictors are then used in a second-step wage equation. Combes and Lafourcade (2011) show that a structural specification encompassing the role of market and supply access in agglomeration economies can also be obtained in a Cournot competition setting.

Unfortunately, structural market and supply access are closely correlated in general, precisely because circular causalities related to agglomeration effects lead households, firms, and intermediate inputs suppliers to choose the same locations.[10] It is therefore difficult to identify their respective effects separately. One also has to keep in mind that the simultaneous presence of knowledge spillovers would suggest adding a standard Harris market potential in the specification, in order to simultaneously take into account pure agglomeration effects coming from the local technological level and labour skills, $A_{c,t}$ and $s_{c,t}$. Nevertheless, it is itself closely correlated with the structural market and supply accesses, and only one of the three variables usually has a significant effect. When structural market access only is considered, one cannot exclude the possibility that it captures agglomeration effects other than those at play in economic geography models à la Dixit and Stiglitz for instance, even if the approach is structural.

3.2 Industrial specialisation and diversity

The theory used to ground the role of location size on local productivity makes it obvious that most effects should be specific to the industry. They depend on structural parameters such as trade and communication costs, the degree of product differentiation, or the degree of increasing returns to scale, which are a priori all specific to the industry. This suggests that, when a reduced form approach is used, heterogeneous effects of density, land area and Harris market potential across industries could be considered, as suggested by Subsection 2.1.2. In other words, the first way of considering the role of local industrial structure is to investigate industry-specific impacts of determinants of urbanisation economies. At the other extreme, theory can be used to construct structural market and supply access variables that are specific to the industry, and which therefore correspond to what is referred to as localisation economies. These are agglomeration effects within the industry, the determinants of which are local characteristics that depend not only on location and date but also industry, the triplet $\{c, s, t\}$ with previous notations.

Usually, authors do not construct structural market and supply access variables that are specific to the industry because necessary data are not available. Alternatively, one can consider in the

[10] Agglomeration economies increase productivity and thus attract firms. This leads to increase the demands of local labour and intermediate inputs as well as wages and input prices, which attract worker and input suppliers. In turn, the inflow of workers and suppliers magnifies productivity gains from agglomeration economies, attracting even more firms, and so on.

specification other variables that characterise the industry within the local economy. One needs to be careful when introducing such variables related to localisation economies in addition to the local economy size variables related to urbanisation economies. Let us first consider the role of the size of the industry within the location. Typically, if all locations had the same share of all industries, the effect of such a variable would not be identified. A location with larger total employment would have more employment in all industries, and higher productivity in an industry could not be attributed more to higher employment in the industry than to higher total employment. Nevertheless, since localisation effects seem to play no role in that case given that all locations have the same industrial composition, one may wish to attribute higher industry productivity in larger cities to higher overall employment in the local economy, i.e. to urbanisation effects. When the industrial share differs across locations for some industries, total and industrial employment are not proportional across locations but one is faced with the same identification issue. Industrial employment can generate productivity gains both when it is higher because total employment at the location is higher, and when the share of the industry is higher for given total employment at the location. These two effects are captured by employment in industry s in location c at date t, $emp_{c,s,t}$, but they can be distinguished by decomposing this employment into the product of its share within the local economy, a variable often labelled specialisation (or concentration in Henderson et al., 1995), and the local size of the economy: $emp_{c,s,t} = spe_{c,s,t}\, emp_{c,t}$, with

$$spe_{c,s,t} = \frac{emp_{c,s,t}}{emp_{c,t}}.$$

To ease interpretation, Combes (2000) argues that in a specification in logarithm, one has to consider total employment (or employment density) next to specialisation. Both these variables are expected to have a positive impact.

Because all variables are in logarithm, the same parameters would also be identified if total employment (or density) and industrial employment (not specialisation) were considered. However, one needs again to be careful with interpretations. We have:

$$\beta \ln emp_{c,t} + \vartheta \ln spe_{c,s,t} = \varrho \ln emp_{c,t} + \vartheta \ln emp_{c,s,t} \;\text{ with }\; \varrho = \beta - \vartheta. \tag{36}$$

This equation shows that, whereas the effect of specialisation for a given total employment and the effect of industrial employment for a given total employment take the same value ϑ, the effect of total employment for given industrial employment ϱ is equal to the difference between the effect of total employment for a given specialisation β and the effect of industrial employment ϑ. A non-significant estimate for ϱ, as obtained for instance by Martin et al. (2011) for France, does not imply that there is no urbanisation effect, but rather means that the effects of specialisation and total employment, which are usually both positive, compensate.[11] Finally, note that one could

[11] Earlier contributions by Glaeser et al. (1992) and Henderson et al. (1995) also consider the share and not the level

consider the density of industrial employment (rather than its level), as we considered the density of total employment and not its level. We do not advise using this specification as it can lead to the same possible mis-interpretations as for the industrial employment level.

Jacobs (1969) made popular the intuition that industrial diversity could be favourable as there could be cross-fertilisation of ideas and transmission of innovations between industries. This has been for instance formalised by Duranton and Puga (2001) and many summary measures of diversity have been proposed. The most used is probably the inverse of a Herfindahl index constructed from the shares of industries within local employment:

$$div_{c,t} = \left[\sum_s \left(\frac{emp_{c,s,t}}{emp_{c,t}} \right)^2 \right]^{-1}.$$

Since specialisation is also introduced in the specification, interpretation is easier if one removes the own industry from the computation of $div_{c,t}$. In that case, whereas specialisation relates to the role of the industry local share, diversity relates the role of the distribution of employment over all other industries, and the two indices clearly capture two different types of mechanisms. In particular, whereas specialisation is a determinant of localisation economies, the Herfindahl index is a determinant of urbanisation economies. Note that when the number of industries is large, it makes little difference to drop the own industry from computations, and the correlation between the Herfindahl indices obtained with and without the own industry is large.

The Herfindahl index has the bad property of taking values largely influenced by the number of units, industries here, from which it is computed. The range of variations of $div_{c,t}$ is $[1, S_{c,t}]$, where $S_{c,t}$ is the total number of industries active in location c at date t. When detailed industrial classifications are used, $S_{c,t}$ can vary a lot across locations and the Herfindahl index reflects this number more than the actual distribution of employment between industries. For this reason, Combes et al. (2004) propose assessing the role of industrial diversity by introducing the Herfindahl index in regressions simultaneously with the number of locally active industries meant to capture the unevenness of the distribution of industries over space.

Another solution consists of moving to other types of industrial diversity indices, keeping in mind that all have weaknesses. For example, some authors propose using the so-called Krugman index introduced by Krugman (1991b). The index is sometimes called the Krugman specialisation index, which is misleading since it actually measures an absence of diversity and specialisation refers to another concept as we have just seen. The Krugman index is a measure of distance

of industrial employment to capture localisation economies. However, because these authors study the determinants of industrial employment growth, and not the productivity level, they argue that the level of industrial employment must be simultaneously introduced, and its effect is identified because not all variables are expressed in logarithm. In that case, identification is granted only thanks to non-linearities, and results can be misleading as emphasised by Combes (2000). We return to this point in Section 6.1.

between the distributions of industry shares in the location and at the global level:

$$K\text{-}Index_{c,t} = \sum_s \left| \frac{emp_{c,s,t}}{emp_{c,t}} - \frac{emp_{s,t}}{emp_t} \right|,$$

where $emp_{s,t}$ is employment in industry s at the global level and emp_t is total employment.

As the Krugman index can take the value zero, it is not possible to express it in a logarithm form. A diversity index can be constructed as the logarithm of one minus the Krugman index. Note that here diversity is maximal when the local distribution of employment across industries is identical to the global one, while an equal share of employment across all sectors at the local level corresponds to a less diverse situation.

Instead of using own-industry specialisation and diversity variables in a specification, one could introduce a full set of variables corresponding to specialisation in all industries. The coefficients of these variables could depend both on own industry and the industry for which specialisation is computed, so that one ends up with a matrix of coefficients. This way one could identify local externalities within each industry and externalities between any two industries (which would not be constrained to be symmetrical). This would possibly correspond more to what Jacobs (1969) had in mind when she said that a number of other industries have a positive effect on own productivity but certainly not all of them as the diversity indices implicitly assume. The effect of specialisation at distant locations could also be assessed by introducing some Harris market potential variables constructed using industrial employment. However, there may be a lack of variation in the data to identify all the effects in these alternative specifications. Endogeneity issues are also magnified, as explained in more detail in Subsection 4.2. All variables should be instrumented at the same time and this can prove to be very difficult in practice.

Finally, for given local total and industrial employment, another industrial characteristic that may influence the magnitude of localisation economies is whether local industrial employment is concentrated in a small number of firms or evenly split among many firms. Typically large firms could be more able to internalise some of the local effects while small firms would have more difficulty avoiding outgoing knowledge spillovers but could also simultaneously benefit more from spillovers. The local distribution of firm sizes also influences the degree of competition on local input markets and on local non-tradable good markets. With this type of intuition in mind, Glaeser et al. (1992) suggest considering the average firm size within the local industry (in fact they consider its inverse) as an additional determinant of localisation economies:

$$size_{c,s,t} = \frac{emp_{c,s,t}}{n_{c,s,t}}.$$

where $n_{c,s,t}$ is the number of firms in industry s in location c at time t. This variable can also be considered simultaneously with a Herfindahl index computed using the shares of firms within local

industrial employment as proposed by Combes et al. (2004). This index captures local productive concentration and can be written as:

$$pcon_{c,s,t} = \sum_{j \in \{c,s,t\}} \left(\frac{emp_{j,t}}{emp_{c,s,t}} \right)^2,$$

where $emp_{j,t}$ is the employment of plant j. Note that the range of variations of this variable depends on the number of plants active in the local industry $n_{c,s,t}$, and this number thus needs to be introduced simultaneously in the specification. Alternatively and more intuitively, one may prefer to introduce instead the average firm size, $size_{c,s,t}$ (as, when expressed in logarithm, $spe_{c,s,t}$, $size_{c,s,t}$, and $n_{c,s,t}$ are collinear).

Importantly, as $size_{c,s,t}$ and $pcon_{c,s,t}$ depend on the location choices of firms and their scale of production, which are directly influenced by the dependent variable (local productivity), their use leads to more serious endogeneity concerns than the one of other local characteristics. These concerns are discussed in more detail in Section 4. Absent a solid instrumentation strategy, one should avoid introducing these determinants of localisation economies in the specification.

3.3 Human capital externalities

Another strand of the literature has tried to identify human capital externalities. Local productivity is regressed on an indicator of local human capital, typically the share of skilled workers in local employment or the local ratio between the numbers of skilled workers and unskilled workers. Somewhat surprisingly, other local characteristics capturing agglomeration effects are most often not introduced simultaneously in the regressions except in a few cases such as Combes et al. (2008a). There is no underlying theoretical reason as we saw that the various agglomeration economy channels may depend on all local characteristics. Furthermore, the human capital variable may be correlated with local characteristics which are not controlled for, such as density with which it is usually positively correlated, and therefore it does not capture the effect of human capital only.

Another difficulty arises from the fact that, beyond some human capital externalities, the estimated coefficient for the local share of skilled workers captures the imperfect substitutability between skilled and unskilled workers. When this share increases, both types of workers can benefit from the externalities but unskilled workers benefit from an extra positive effect due to the fact that they become relatively less numerous, which increases their marginal productivity. Conversely, skilled workers are negatively affected by this substitution effect. We illustrate this identification issue by considering the following local production function that extends our previous framework:

$$y_{c,t} = \left[\left(A^H_{c,t} H_{c,t} \right)^\rho + \left(A^L_{c,t} L_{c,t} \right)^\rho \right]^{\frac{\alpha}{\rho}} K^{1-\alpha}_{c,t} \tag{37}$$

where $A^j_{c,t}$ is the productivity of workers with skills j with $j = H$ for high-skill workers and $j = L$ for low-skill workers, $H_{c,t}$ is the number of high-skill workers, $L_{c,t}$ is the number of low-skill workers, and ρ is a parameter such that $\rho < 1$. The production function is Cobb-Douglas in labour and other inputs, $K_{c,t}$, and the labour component is a CES function in high- and low-skill workers with an elasticity of substitution equal to $-1/(1-\rho)$. As previously, workers are counted in terms of efficient units such that:

$$H_{c,t} = \sum_{i \text{ high-skilled } \in \{c,t\}} s_{i,t}\ell_{i,t} \tag{38}$$

$$L_{c,t} = \sum_{i \text{ low-skilled } \in \{c,t\}} s_{i,t}\ell_{i,t} \tag{39}$$

with $\ell_{i,t}$ the number of hours worked and $s_{i,t}$ the number of efficient labour units per hour of individual i at date t. As regards the human capital externality, the ratio between the numbers of high-skill and low-skill workers $S_{c,t} = H_{c,t}/L_{c,t}$ is supposed to influence the productivity of workers differently depending on their skills such that:

$$A^H_{c,t} = (S_{c,t})^{\gamma^H} \quad \text{and} \quad A^L_{c,t} = (S_{c,t})^{\gamma^L}. \tag{40}$$

where γ^j captures the magnitude of human capital externalities for workers with skills j. For simplicity's sake, we assume here that $S_{c,t}$ does not affect any other agglomeration channel, namely the prices of output and other inputs, and that no other local characteristic plays a role. It is possible to solve for wages at the individual level in the same way we did in Section 2 using first-order conditions to determine the optimal use of labour and capital. The wages of high- and low-skill workers, $w^j_{i,t}$ for $j = H, L$, is obtained as:

$$w^H_{i,t} = \frac{\alpha}{(1-\alpha)^{1-1/\rho}} r_{c,t}^{1-1/\rho} p_{c,t}^{1/\alpha} s_{i,t} \left(A^H_{c,t}\right)^\rho \left[\left(A^H_{c,t}\right)^\rho + \left(A^L_{c,t}\right)^\rho S_{c,t}^{-\rho}\right]^{\frac{1-\rho}{\rho}}, \tag{41}$$

$$w^L_{i,t} = \frac{\alpha}{(1-\alpha)^{1-1/\rho}} r_{c,t}^{1-1/\rho} p_{c,t}^{1/\alpha} s_{i,t} \left(A^L_{c,t}\right)^\rho \left[\left(A^H_{c,t}\right)^\rho + \left(A^L_{c,t}\right)^\rho S_{c,t}^{-\rho}\right]^{\frac{1-\rho}{\rho}} S_{c,t}^{1-\rho}. \tag{42}$$

The wage elasticities with respect to $S_{c,t}$ for high- and low-skill workers respectively can be derived as:

$$\delta^H_{c,t} = \gamma^H - \phi_{c,t}(1-\rho)\left(1 + \gamma^H - \gamma^L\right) \tag{43}$$

$$\delta^L_{c,t} = \gamma^L + (1-\phi_{c,t})(1-\rho)\left(1 + \gamma^H - \gamma^L\right) \tag{44}$$

where $\phi_{c,t}$ is the ratio between the wage bill of high-skill workers and the total wage bill.

Several comments can be made about these elasticities. Most importantly, they do not capture

the effect of human capital externalities only but also the degree of substitution between high-skill and low-skill workers. Suppose that human capital externalities are present for both types of workers but their impact is greater on high-skill workers than on low-skill workers, $\gamma^H > \gamma^L$. In that case, the wage elasticity for low-skill workers with respect to $S_{c,t}$, $\delta^L_{c,t}$, is always positive as both the externality and the substitution effects increase their productivity. By contrast, the wage elasticity for high-skill workers, $\delta^H_{c,t}$, may be either positive or negative, as the substitution effect goes in the opposite direction from the externality effect. As acknowledged by Moretti (2004b) and Ciccone and Peri (2006), the magnitude of human capital externalities cannot be recovered from simple regressions of the logarithm of wage on $S_{c,t}$, even conducted separately for low-skill and high-skill workers. However the specification can be easily augmented to identify both externality and substitution effects.

Wage elasticities $\delta^H_{c,t}$ and $\delta^L_{c,t}$ in (43) and (44) vary across locations since there is no reason the wage bill ratio $\phi_{c,t}$ should be constant over space. This suggests regressing the logarithm of wage not only on the human capital variable $S_{c,t}$ but also on its interaction with $\phi_{c,t}$ (while also including in the specification individual fixed effects, individual variables, and local variables affecting other types of agglomeration economies). Regressions should be run separately for high-skill and low-skill workers as the coefficients for the two variables are not identical for the two types of workers. According to (43) and (44), one recovers four coefficients that can be used to estimate the three parameters γ^H, γ^L and ρ. The model is over-identified, which makes it possible to conduct a specification test.

An alternative approach has been proposed by Ciccone and Peri (2006) but only the average effect of human capital externalities can be recovered and not those specific to each type of workers. We present this approach in a simplified way. The authors first compute a local average wage weighted by the share of each worker type in local employment, $w_{c,t} = s_{c,t} w^H_{c,t} + (1 - s_{c,t})\, w^L_{c,t}$, with $s_{c,t}$ the share of high-skill workers in local employment. The elasticity of this average wage with respect to $S_{c,t}$, holding $s_{c,t}$ constant, is given by:

$$\frac{\partial \log w_{c,t}}{\partial \log S_{c,t}} = \phi_{c,t}\gamma^H + (1 - \phi_{c,t})\, \gamma^L. \tag{45}$$

This relationship is strictly valid for variations over time in the short run in line with the definition of the elasticity. The authors make the approximation that it can be used to study long-run variations of the logarithm of wage between two dates t and t' (1970 and 1990 in their application) when the logarithm of $S_{c,t}$ varies holding constant the local share of workers. More precisely, they first construct a city wage index at date t' considering the local composition of workers at date t:

$$\widetilde{w}_{c,t'} = s_{c,t} w^H_{c,t'} + (1 - s_{c,t})\, w^L_{c,t'} \tag{46}$$

The log-wage difference $\log \widetilde{w}_{c,t'} - \log w_{c,t}$ is then regressed on $\log S_{c,t'} - \log S_{c,t}$ to recover an effect

supposed to be the weighted average of the effects of human capital externalities given by (45).

What remains unclear is the source of variations over time of $S_{c,t}$. Holding the share of high-skill workers in total employment $s_{c,t}$ constant implies that the ratio between the numbers of high-skill and low-skill workers, $S_{c,t}$, is constant too. Another issue arises because the right-hand side of (45) is considered to be a constant coefficient whereas it clearly varies across cities since $\phi_{c,t}$ is specific to the city. Finally, even if the wage $\widetilde{w}_{c,t'}$ is supposed to be computed fixing the local composition of workers to its value at date t, its computation involves the wages of both skill groups at date t', $w^j_{c,t'}$. These are not the wages that workers would have had when holding constant the composition of employment. Indeed the actual variation of wages between the two dates may have been influenced by the changes in the local composition of workers.

The use of a CES production function emphasizes the role of the elasticity of substitution between high-skill and low-skill workers, which can be recovered from the estimations. It is possible to conduct a similar analysis with a Cobb-Douglas production function although the elasticity of substitution is then fixed and equal to -1 (in particular, we get a Cobb-Douglas specification in our setting when ρ tends to zero). In that case, local labour cost shares are constant and they are given by the Cobb-Douglas coefficients of the two groups. Nevertheless, the procedure we propose can still be applied if the Cobb-Douglas technology is allowed to differ across locations.

Finally, alternative variables can be considered to measure local human capital externalities, such as the share of high-skill workers in total employment. The choice of a variable ultimately relies on the choice of an ad hoc functional form. For instance, Moretti (2004b) and Combes et al. (2008a) regress the logarithm of individual wages on the local share of high-skill workers in total employment controlling for an individual fixed effect, as well as individual and local characteristics. Even when the specification is estimated separately for high-skill and low-skill workers, the issue remains that only a composite of the externality effect and the substitution effect is identified. To go further and identify separately the two effects, it might be worth augmenting the specifications with the interaction of the human capital variable and the local share of high-skill workers in the wage bill.

4 Estimation strategy

Now that the links between theory and empirical specifications, as well as the interpretation of estimated coefficients, have been clarified, we move to a number of empirical issues. First, we discuss the use of TFP rather than nominal wage as a measure of productivity. We then turn to endogeneity issues which emerge when estimating wage or TFP specifications. We present the solutions proposed in the literature to deal with these issues as well as their limits. We finally discuss a series of other empirical issues regarding spatial scale, functional forms, observed skills measures, and spatial lag models.

4.1 Wages versus TFP

So far, we have mostly considered nominal wage at the worker level as our measure of productivity. Alternatively, one may wish to use a measure at the firm level such as output value or value added. It is possible to derive a specification for such a measure which is consistent with the production function used in Section 2. Let us rewrite the production function at the firm level as:

$$Y_{j,t} = \frac{A_{c,t}}{\alpha^{\alpha}(1-\alpha)^{1-\alpha}} (s_{j,t}L_{j,t})^{\alpha} K_{j,t}^{1-\alpha}, \tag{47}$$

where j denotes the firm, $Y_{j,t}$ is the firm output, $s_{j,t}$ corresponds to average labour skills, which are allowed to vary across firms, $L_{j,t}$ and $K_{j,t}$ are labour and other inputs, respectively, while $A_{c,t}$ is the technological level supposed to be local (we could alternatively consider that it varies across firms within the same local labour market but this does not change the reasoning and we prefer to stick to a simple specification). Output value is given by $p_{j,t}Y_{j,t}$ where $p_{j,t}$ is the average income of the firm per unit produced (see Footnote 1 for more details). The logarithm of TFP can be recovered as:

$$\ln p_{j,t}Y_{j,t} - \alpha \ln L_{j,t} - (1-\alpha)\ln K_{j,t} = \ln \frac{p_{j,t}A_{c,t}s_{j,t}^{\alpha}}{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}. \tag{48}$$

Equation (48) for TFP is equivalent to (3) in logarithm for wage. It can be used to relate the logarithm of TFP (rather than wage) to some local characteristics, density among others, which determine the channels through which agglomeration economies operate, i.e. the firm price $p_{j,t}$, average labour skills $s_{j,t}$, and the local technological level $A_{c,t}$.

If value added is reported in the data set instead of output value, intermediate consumption can be taken into account in the production function. For instance, consider that production is Leontieff in intermediate consumption denoted $I_{j,t}$ with share in output a and the Cobb-Douglas function (47):

$$Y_{j,t} = Min\left(\frac{I_{j,t}}{a}, \frac{A_{c,t}}{\alpha^{\alpha}(1-\alpha)^{1-\alpha}} (s_{j,t}L_{j,t})^{\alpha} K_{j,t}^{1-\alpha}\right). \tag{49}$$

Profit maximisation yields that intermediate consumption is proportional to production, and this leads to:

$$\ln(p_{j,t}Y_{j,t} - \nu_{j,t}I_{j,t}) - \alpha \ln L_{j,t} - (1-\alpha)\ln K_{j,t} = \ln \frac{(p_{j,t} - a\nu_{j,t})A_{c,t}s_{j,t}^{\alpha}}{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}. \tag{50}$$

where the left-hand side is TFP measured now in terms of value added, with $\nu_{j,t}$ the unit price of intermediate input. This makes it possible to conduct the analysis in a similar way as when TFP is measured in output value. The interpretation of estimated parameters is slightly different since the output price is now net of the unit cost of intermediate consumption.

There are two important differences with a wage analysis, which arise because the term that depends on local characteristics is $p_{j,t}A_{c,t}s_{j,t}^{\alpha}$ when one considers TFP in output value, whereas

it was $\left(p_{c,t}A_{c,t}/(r_{c,t})^{1-\alpha}\right)^{1/\alpha} s_{c,t}$ in the case of the nominal wage (see equation 3). The local cost of inputs other than labour does not enter the expression of output value and the determinants of agglomeration economies only capture effects related to technological level, output price, and average skills. This means that land and housing prices no longer play a role. This is clearly an advantage since we saw that the interpretation of the effect of housing price is difficult for wage regressions and the use of this price as an explanatory variable raises serious endogeneity concerns. Moreover, the elasticity of agglomeration economies obtained from TFP regressions must be multiplied by one over the share of labour in the production function $1/\alpha$ to be directly comparable with the one obtained from wage regressions. For these two reasons, the economic interpretation of the impact of local characteristics is not the same when studying TFP or wage.

It is also important to note that wages are usually only proportional and not equal to labour productivity by a factor that depends on the local monopsony power of the firm. This proportionality factor may be correlated with some local determinants of agglomeration economies but one may wish to avoid considering its spatial variations as part of agglomeration effects. This may be the case when differences in local monopsony power result from differences in institutional features, which occur for instance between countries, and not from differences in the degree of competition on local labour markets. The use of TFP avoids making any assumption about the relationship between the local monopsony power and agglomeration economies. Finally, note that in the framework proposed here, agglomeration effects may operate at the firm level and not only at the local level as in previous sections, since the output price $p_{j,t}$ and average labour skills $s_{j,t}$ are now specific to the firm. This may also be considered for wages, but we postpone the related discussion until Section 4.4.

Empirically, an additional concern is that firm TFP, the left-hand side in (48), is not directly observable in data sets and computing its value requires estimating parameter α.[12] However, output, labour, and other inputs are simultaneously determined by the firm, which causes an endogeneity issue that can potentially bias the estimated coefficient obtained from ordinary least squares. Several methods have been proposed to estimate α consistently, such as a Generalised Method of Moments (GMM) approach applied to the specification of output value in first difference (to deal with firm unobservables) using lagged values of labour and other inputs as instruments in the spirit of Arellano and Bond (1991) and followers, or sophisticated semi-parametric approaches to control for unobservables which make use of additional information on investment Olley and Pakes (1996) or intermediate consumption (Levinsohn and Petrin, 2003). There is no consensus on a method that would be completely convincing and robustness checks have to be conducted using several alternative approaches.

Moreover, agglomeration variables may be endogenous too for the reasons we develop in the

[12] One can relax the assumption of constant returns to scale and also estimate parameters for inputs other than labour without requiring that their total share in input costs is equal to $1-\alpha$.

next subsection, and this issue needs to be addressed. One way to proceed consists of applying a two-stage approach where the production function is estimated in a first stage with one of the alternative methods we have just cited and no local variable is introduced. Local-time averages of residuals are then computed and regressed in a second-stage on some local characteristics. We detail below approaches to deal with the endogeneity of local characteristics in second stage. Alternatively, local-time fixed effects can be introduced in a first stage and their estimators regressed in a second stage, in the spirit of what was proposed for individual wages (see Combes et al., 2010, for more details). This second approach has the advantage of properly controlling at the individual level for unobserved local shocks that may be correlated with firm variables. A last approach consists in estimating a specification of output value $p_{j,t}Y_{j,t}$ including both inputs and local characteristics as explanatory variables, instrumenting variables all at once. This was for instance proposed by Henderson (2003) who estimates an output value specification with the generalized method of moments.

4.2 Endogeneity issues

We now detail the various endogeneity problems that can occur and approaches that have been proposed to solve them. When estimating the effect of local characteristics on individual outcome, endogeneity can occur both at the individual level and at the local economy level. To see this, we rewrite equation (6) as:

$$y_{i,t} = u_i + X_{i,t}\theta + \sum_c \left[Z_{c,t}\gamma + \eta_{c,t}\right] 1_{\{c(i,t)=c\}} + \epsilon_{i,t} \tag{51}$$

where $1_{\{c(i,t)=c\}}$ is a dummy variable equal to 1 when individual i locates in c at date t. This expression involves local effects related to observables, $Z_{c,t}$, and unobservables, $\eta_{c,t}$, on every local markets, and makes explicit the location choice $1_{\{c(i,t)=c\}}$ which is made at the individual level.

There is an endogeneity issue at the local level when a variable in $Z_{c,t}$, density for instance, is correlated with the local random component $\eta_{c,t}$. This can happen because of reverse causality or the existence of some missing local variables that affect directly both density and wages. Reverse causality is an issue when higher local average wages attract workers, as this increases the quantity of local labour and thus density. In that case, one expects a positive bias in the estimated coefficient of density (provided that density has a positive effect on wages due to agglomeration economies).

There is a missing variable problem when for instance some local amenities not included in $Z_{c,t}$ are captured by the local random term and they determine both local density and wages. Productive amenities such as airports, transport infrastructures and universities, increase productivity and attract workers, which makes the density increase. In that case, a positive bias in the estimated coefficient of density is also expected. In line with Roback (1982), consumption amenities such as cultural heritage or social life increase the attractiveness of some locations for workers

and thus make density higher. Such amenities do not have any direct effect on productivity but the increase in housing demand they induce makes land more expensive. As a result, local firms use less land relatively to labour, and this decreases labour productivity when land and labour are imperfect substitutes. This causes a negative bias in the estimated coefficient of density since density is positively correlated with missing variables that decrease productivity.

Finally, the unobserved local term captures among other things the average of individual wage shocks at the local level. This average may depend on density as workers in denser local markets may benefit from better wage offers due for instance to better matching. One may consider that matching effects are part of agglomeration economies and then there is no endogeneity issue. Alternatively, one may be interested solely in the effects of knowledge spillovers and market access for goods captured by density, in which case there is an expected positive bias in the estimated effect of density due to the contamination by matching mechanisms.

Endogeneity concerns can also arise at the individual level when location dummies $1_{\{c(i,t)=c\}}$ are correlated with the individual error term $\epsilon_{i,t}$. This occurs when workers sort across locations according to individual characteristics not controlled for in the specification such as some of their unobserved abilities. We emphasise in Subsection 2.1 the importance of considering individual fixed effects u_i to capture the role of any individual characteristic constant over time. However, workers might still sort across space according to some time varying unobserved characteristics entering $\epsilon_{i,t}$.

Endogeneity at the individual level also emerges when workers' location choices depend on the exact wage that they get in some local markets, typically when they receive job offers associated with known wages. Notice that this type of bias is closely related to matching mechanisms although there is here an individual arbitrage between locations whereas matching effects mentioned earlier rather refer to a better average situation of workers within some local markets. Importantly, as long as individual location decisions depend only on the explanatory terms introduced in the specification, which can go as far as the individual fixed effect, some time-varying individual characteristics such as age, and a location-time fixed effect, there is no endogeneity bias. Combes et al. (2011) details these endogeneity concerns.

4.3 Dealing with endogenous local determinants

The literature has mostly addressed endogeneity issues at the local level using several alternative strategies. A simple approach consists of including time-invariant local fixed effects in specifications estimated on panel data to deal with missing local variables constant over time. Some authors instrument the local determinants of agglomeration economies using additional variables such as local historical or geological variables. Estimations with GMM, where lagged values of local determinants themselves are used for instrumentation, have been considered too but their validity relies on stronger assumptions. Finally, other articles exploit natural experiments involving a

shock on local characteristics related to agglomeration economies. This section examines these various strategies. The reader may also refer to the chapter by Baum-Snow and Ferreira (2015) for additional considerations on causality.

By contrast, we are not aware of non-structural contributions dealing with endogeneity at the individual level, to the extent that some concerns would remain in the most complete specifications including both individual and location-time fixed effects. Structural approaches considering dynamic frameworks like those presented in Section 2.4 are clearly a natural direction to go to consider endogenous individual location choices.

4.3.1 Local fixed effects

One reason why local determinants of agglomeration economies can be endogenous is that some missing variables determine them simultaneously with the local outcome. In particular, this is the case when there are missing amenities that affect both local productivity and local population. A strategy for coping with this issue when having panel data at hand is to include local fixed effects in the estimated specification. There are several reasons why this strategy may not work well. First, it does not deal with missing variables that evolve over time: for instance new airports or stations are built or improved over the years depending precisely on their local demand and the performance of local firms and workers. Second, time invariant local fixed effects do not help in solving the endogeneity issue due to reverse causality, such that higher expected wages or productivity in a location attract more firms and workers. Third, identification relies on time variations of the local outcome and local determinants of agglomeration economies only. If the variations of local determinants are mis-measured, which is likely to happen as local determinants are often computed from samples of limited size and variations are often considered only in the short run because the time span of panels is in general pretty short, estimated effects can be highly biased because of measurement errors. This kind of problem can be particularly important for local characteristics which vary little across time, for instance because the economy is close to a spatial equilibrium.[13] Their effect is difficult to identify separately from the role of permanent characteristics that affect productivity without being related to agglomeration economies. Nevertheless, one can try to identify their effect by using an instrumentation strategy applied to a specification in level.

4.3.2 Instrumentation with historical and geological variables

An alternative strategy for coping with endogeneity at the local level consists of finding instruments that deal with both reverse causality and missing amenities. Instruments should verify two conditions: relevance and exogeneity. Instruments are relevant when they are correlated with the instrumented variables $Z_{c,t}$, and they are exogenous when they are not correlated with the

[13] This does not necessarily mean that they do not shape the magnitude of agglomeration economies.

aggregate random term $\eta_{c,t}$. Two necessary conditions for exogeneity are that instruments are not correlated with missing local variables and not determined by the outcome.

Several sets of instruments have been proposed. The first one consists of historical instruments and more particularly long lagged values of variables measuring agglomeration economies (see Ciccone and Hall, 1996; Combes et al., 2008a). Historical values of population or density are usually considered to be relevant because local housing stock, office buildings and factories last over time and create inertia in local population and economic activity. If lags are long enough (say 150 years), instruments are believed to be exogenous because of changes in the type of economic activities (agriculture to manufacturing then services) and sometimes wars that reshaped the area under study. Local outcomes today are therefore unlikely to be related to components of local outcomes a long time ago that probably affected historical population. However there are local permanent characteristics that may have affected past location choices and still affect local productivity today, such as the centrality of the location in the country, a suitable climate, or geographical features like access to the coast or presence of a large river. If these features are not properly controlled for in regressions, local historical population may not be exogenous.

The second set of instruments consists of geological variables related to the subsoil of the location (see Rosenthal and Strange, 2008; Combes et al., 2010). These variables typically describe soil composition, depth to rock, water capacity, soil erodability, and seismic and landslide hazard. They are believed to be relevant because the characteristics of soils were important for agriculture centuries ago, even millennia ago, and manufacturing and services have then developed where human settlements were already located. They are believed to be exogenous because people may have had only a negligible effect on soil and geology, and these do not influence productivity of most modern activities.

Some authors argue that consumption amenities can be used as instruments since according to the Roback (1982) model, they are relevant because they attract workers and therefore determine local population, and they are exogenous as they would not directly affect local productivity. This is not fully granted, however, because the inflow of workers puts pressure on local land markets, which in turn gives firms incentives to substitute labour for land in the production process, as we have already argued above. As a result, productivity can be affected and consumption amenities are not exogenous. Therefore we advocate using consumption amenities as control variables rather than as instruments when they are available in datasets.

In practice, historical variables are usually found to be extremely relevant instruments, in particular past population, indicating major inertia in the distribution of population over space. Geological variables are also found to be relevant but to a lesser extent and their power to explain instrumented variables is not very high. Exogeneity can only be properly tested by confronting different sets of instruments with each other, under the assumption that at least one set of instruments is valid. Indeed, the Sargan exogeneity test implicitly compares the estimators obtained

with all the alternative combinations of instruments. The test is passed when these estimators are not significantly different from each other. One has to make the assumption that at least one set of instruments is valid such that the instrumental variable estimator obtained with that set of instruments is consistent. Otherwise, the test could be passed with all instruments being invalid and the instrumental variable estimators obtained with the different combinations of instruments all converging to the same wrong value. As an implication, making an exogeneity test using only very similar instruments (for instance population 150, 160 and 180 years ago) is not appropriate since the estimated coefficient could be biased the same way in all cases and the over-identification test would then not reject exogeneity. An over-identification test using different types of instrument which are not of same nature is more meaningful. For instance, it is likely that historical and geological variables satisfy this property: even if geology initially influenced people location choices a very long time ago, many other factors have also determined the distribution of population across space since then and make local historical population a century ago less related to local geology. Some authors, such as Stock and Yogo (2005), have started to develop weak instrument tests that assess whether different instruments have enough explanatory power of their own and can be used together to conduct meaningful over-identification tests. Such tests should be systematically reported.

Lastly, since Imbens and Angrist (1994), it has been emphasised that instrumentation identifies a Local Average Treatment Effect (LATE) only, i.e. an effect specific to the chosen instruments, and not necessarily the average treatment effect (ATE). Some differences between the two occur when instruments differently weight observations, locations here, in regressions. For instance, current total population may be instrumented with historical urban population rather than historical total population because of data availability issues (see Combes et al., 2008a). In that case, the instrument is more relevant for locations with a current population which is large. Indeed, the instrument takes the value zero for all locations with no urban population a long time ago, and still varies with positive value for locations of large size with positive urban population a while ago. Overall, this argues for considering different sets of instruments, testing whether they lead to similar estimates as mentioned earlier, and keeping in mind the arguments developed here for the interpretation of different estimates.

4.3.3 Generalised Method of Moments

A third strategy that has been used to cope with endogeneity issues when having panel data is to use a GMM approach to estimate the specification in first difference while using lagged values of variables as instruments, both in level and first difference. Two main types of specification involving determinants of agglomeration economies have been estimated that way: dynamic specifications of employment at the city-industry level (Henderson, 1997; Combes et al., 2004) and static or dynamic specifications of TFP or wages (Henderson, 2003; Mion, 2004; Graham et al., 2010; Martin et al.,

2011). As detailed in Subsection 4.1, articles on productivity typically specify in logarithm the firm production or value added as a function of labour, other inputs (usually physical capital), local variables determining agglomeration economies, possibly earlier in time, and a firm fixed effect capturing time-invariant firm and local effects. The specification is rewritten in first difference between t and $t-1$ to eliminate the firm fixed effect. A similar strategy is implemented at the local level when no firm level data are available. When the effects of all variables are estimated in a single step, first differences of labour, capital and local variables are simultaneously instrumented by their past values in $t-k$ with $k \geq 2$, and/or by their past levels. When a two-step strategy is implemented such that a TFP specification is first estimated and then either local-time averages of residuals or local-time fixed effects are regressed on local characteristics in a second step, the same kind of instrumentation can be implemented at each step. Lastly, an alternative approach has been proposed by Graham et al. (2010) who specify a VAR model where the first equation relates current labour productivity to its past values and those of local characteristics, and additional equations relate current values of local characteristics to their past values and those of productivity. All equations are simultaneously estimated with dynamic GMM and Granger tests are used to assess the presence of reverse causality between productivity and local characteristics.

As detailed in Section 6.1 below, studies of employment dynamics specify city-industry employment at time t as a function of its lags at times $t-1$, ..., $t-k$ with $k \geq 1$, other time-varying local characteristics and a city-industry fixed effect. Lags of the dependent variable capture both mean-reversion and agglomeration size effects as argued by Combes et al. (2004), while local characteristics capture other types of agglomeration economies.[14] Again the specification is rewritten in first difference between t and $t-1$, and first-differenced lags of city-industry population are instrumented with past levels before $t-k$ with $k \geq 3$, and other local variables with their value in $t-2$.

The approach is valid when the two conditions of relevance and exogeneity of instruments are verified. Relevance of instruments is usually not an issue as there is some inertia in local variables and the time span is usually short (a couple of decades at most). Exogeneity can be the most problematic issue. Take the example of city-industry employment $y_{z,s,t}$ written in first difference $\Delta y_{z,s,t} = y_{z,s,t} - y_{z,s,t-1}$ and regressed on its lagged value $\Delta y_{z,s,t-1}$. The practice consists in instrumenting $\Delta y_{z,s,t-1}$ with the past level $y_{z,s,t-2}$. Typically, the exogeneity condition is not verified if the shock in the outcome specification, say $\nu_{z,s,t}$, is serially correlated. This causes the shock in first difference $\Delta\nu_{z,s,t}$ to correlate with the employment past level $y_{z,s,t-2}$. Industry-city shocks probably last several years and the exogeneity condition is thus unlikely to hold. One may wish to use as instruments more remote past levels $y_{z,s,t-k}$ with k much larger than two to attenuate the bias but this strategy will also probably fail when the data span 15 or 20 years only. A common practice for testing the validity of the exogeneity condition is to use several lags of the outcome

[14] Note that there are specific interpretation issues that are discussed in Section 6.1.

before $t-1$ as instruments and conduct a Sargan exogeneity test. This practice is dubious since the test relies on instruments all from the same source, the dependent variable itself. As suggested earlier, variables of a different kind should be used as instruments together with past values of the outcome for the test to be meaningful. Overall, we advise against relying on approaches based on GMM with lagged values as instruments to identify the role of local determinants on local outcomes.

4.3.4 Natural experiments

Another strategy for dealing with an endogenous local determinant consists of exploiting the context of a natural experiment that has induced a sizeable localised shock on that determinant which is not directly related to the outcome variable. The general idea of the approach is to evaluate the effect of the variable from the comparison of the average variation in outcome in places which have experienced the shock with the average variation in outcome in comparable places which have not. Sometimes, the quantitative value of the shock is not known, and only its effect (i.e. the change in the agglomeration determinant times the coefficient variable) is identified. To see this, consider the aggregate model:

$$\beta_{c,t} = Z_{c,t}\gamma + \theta_c + \eta_{c,t} \tag{52}$$

where $\beta_{c,t}$ is a local outcome such as a location-time fixed effect estimated in first step on individual data, $Z_{c,t}$ includes the local characteristics that determine agglomeration effects, and θ_c is a location fixed effect capturing among others the role of local time invariant characteristics. A common practice is to make the city fixed effect disappear by rewriting the model in first difference:

$$\Delta\beta_{c,t} = \Delta Z_{c,t}\gamma + \Delta\eta_{c,t} \tag{53}$$

Beyond the fact that controlling for time invariant local effects can raise measurement issues as discussed above, another problem is that the variation in local variable $\Delta Z_{c,t}$ may be correlated with the variation in residual $\Delta\eta_{c,t}$ because of unobserved time-varying amenities or reverse causality. This problem can be circumvented in the case of a natural experiment. Consider that there is a subset denoted *tr* (for *treated*) of N_{tr} locations experiencing a shock, or "treatment", that affects the local variable from date τ onwards such that $Z_{c,t} = \bar{Z}_{c,t} + \phi\, 1_{\{t\geq\tau\}}$ where $\bar{Z}_{c,t}$ is the value of the local variable in the absence of shock, and $1_{\{t\geq\tau\}}$ is a dummy for being affected by the shock. Consider also that there is a subset denoted *ntr* (for *non-treated*) of N_{ntr} locations that do not experience any shock from date τ onwards. The difference-in-differences estimator of the effect of the shock between dates $\tau-1$ and τ is the difference between the average outcomes of the treated

and non-treated locations, given by:

$$\widehat{\phi\,\gamma} = \frac{1}{N_{tr}} \sum_{c \in tr} \Delta\beta_{c,\tau} - \frac{1}{N_{ntr}} \sum_{c \in ntr} \Delta\beta_{c,\tau} \tag{54}$$

This estimator converges to the true effect of the shock $\phi\,\gamma$ provided that the numbers of locations in the treated and non-treated groups tend to infinity and that there is similarity between treated and non-treated locations in terms of growth of local variables and shocks in the absence of treatment:

$$E\left[\Delta\bar{Z}_{c,t} \middle| c \in tr\right] = E\left[\Delta\bar{Z}_{c,t} \middle| c \in ntr\right] \quad \text{and} \quad E\left[\Delta\eta_{c,t} \middle| c \in tr\right] = E\left[\Delta\eta_{c,t} \middle| c \in ntr\right]. \tag{55}$$

Note that when the value of the shock ϕ is observed, it is then possible to recover the marginal impact of the local variable, γ.

The challenge when using a natural experiment is to find a control group which is similar to the treated group such that locations in the two groups would have experienced similar variations in local characteristics absent the shock and such that their unobserved characteristics would have evolved similarly. If this is not the case, strategies based on matching can lead to further comparability between the two groups or regression discontinuity approaches can be used to identify the effect of treatment locally.

A limitation when exploiting a natural experiment, in particular when using these two complementary strategies, is that external validity is not granted. The shock may be specific to a particular context and locations in the treated and non-treated groups may not be representative of the overall set of cities. Therefore the estimator obtained from the natural experiment may not correspond to the average effect of the shock for the whole set of cities.

Some papers such as Hanson (1997), Redding and Sturm (2008), and Greenstone et al. (2010) have achieved some success in using natural experiments when studying the effect of local determinants of agglomeration economies on outcomes of firms. We detail their strategies and conclusions in Section 5.4 concerning the results obtained in the literature.

4.4 Tackling the role of firm characteristics

We have so far considered a production function where the total factor productivity of firms is influenced by location but not by any intrinsic characteristic of firms. It is possible to argue though that firms differ in their management teams, with some being more efficient than others, and this creates some heterogeneity in productivity. Moreover, there can be some sorting of firms across space depending on management efficiency, for instance with firms with the better management teams being created in larger locations. International trade models with heterogeneous firms also imply that only the most able firms can survive in larger markets (see for instance Melitz and Ottaviano, 2008) due to competition effects that are not related to any kind of agglomeration

gains. If such firm selection effects exist and firm heterogeneity is not properly taken into account, estimated effects of local characteristics such as city size are biased.

Heterogeneity in firm productivity can be taken into account in specifications of firm output value derived in Subsection 4.1 by making the TFP specific to the firm rather than to the area in the same way we did for output and input prices. A possible way of taking into account firm heterogeneity in wage regressions is to include firm fixed effects in wage specifications such as (6) which becomes:

$$y_{i,t} = u_i + v_{j(i)} + X_{i,t}\theta + Z_{c(i,t),t}\gamma + \eta_{c(i,t),t} + \epsilon_{i,t}. \tag{56}$$

where $j(i)$ is the firm of individual i and v_j is a firm fixed effect. Two estimation issues need to be discussed. First, it is never possible to control properly for all productive amenities by including explanatory variables at the local level in the regression. Firm fixed effects are thus bound to capture the effect of any omitted local variable not varying over time and they thus cannot simply be interpreted as firm effects. From a theoretical point of view, this is crucial when trying to interpret the correlation between worker and firm fixed effects. This correlation does not necessarily capture the effect of a worker-firm match but could also capture the effect of a worker-area match with some sorting of firms depending on unobserved local characteristics. Nevertheless, firm fixed effects can still be considered as controls.

Second, it is difficult, if not impossible, to take into account time-varying local unobservables in the computation of standard errors. Indeed, the two-step approach proposed in Subsection 2.1.1 cannot be applied since local-time fixed effects cannot be identified separately from firm fixed effects. This occurs because firms do not move across space and the local average of their effects is then confounded with local effects. The larger the unobserved local effects, the larger the possible bias in standard errors derived from least square estimation. Some determinants of agglomeration economies could appear to have a significant effect, whereas they would not if unobserved local effects were properly considered.

An alternative approach consists of introducing proxies in the specification for firm characteristics related for instance to management or organisation, instead of firm fixed effects. One can then apply the two-stage approach to properly take into account local unobservables in the computation of standard errors. Such proxies are hard to find however and, when estimations are conducted in a single step, firm variables may also capture their effects which can be due to agglomeration economies. In particular, some authors use firm size as a regressor and do not control for local-time fixed effects (see for instance Mion and Naticchioni, 2009). Firm size may not only capture firm productivity but also agglomeration gains from increasing returns to scale due to a better market access. One may try to distinguish firm productivity by rather using firm size centred with respect to its local average. Another clear limitation to controlling for firm size is that it depends on time-dependent shocks that also affect wages. This causes a simultaneity bias in the estimations. Note that all these issues are common to most firm observed characteristics.

Firm heterogeneity can itself be used to distinguish agglomeration effects from competition effects as proposed by Combes et al. (2012b). This paper considers a value-added specification where only labour, capital and skills are introduced. Firm TFP is measured with the residual computed at the firm level. An economic geography model with heterogeneous firms shows that a test for the presence of agglomeration and competition effects can then be conducted by comparing firms' TFP distributions in small and large cities. If the distribution in large cities is a right-shifted version of the distribution in small cities, all firms in large cities benefit from agglomeration effects. If the distribution in large cities is rather a left-truncated version of the distribution in small cities, competition is fiercer in large cities which leads to a larger share of the least productive firms being unable to survive there. Estimations from French data taking into account both the right-shift and left-truncation transformations support the presence of agglomeration effects but not the presence of competition effects.

4.5 Other empirical issues

4.5.1 Spatial scale

Papers differ in the spatial scale at which the impact of local determinants is measured. There are two main reasons for that: there is no real consensus on the spatial scope at which each agglomeration mechanism takes place and any local determinant captures in general several mechanisms, the relative intensity of which can differ across spatial scales. Theory makes it clear that the spatial scope of agglomeration effects depends on their type. For instance, whereas technological spillovers often require face-to-face contacts, other agglomeration effects such as input-output linkages could take place at a larger scale such as the region. The issue is in fact more complicated as changing the size of the spatial units usually involves changing their shape, and both changes create Modifiable Areal Unit Problems (MAUPs) that were already mentioned above. However, Briant et al. (2010) show in the particular case of the effect of local density on individual wages, that changing shapes is of secondary importance for the estimates compared to taking into account individual unobserved heterogeneity with individual fixed effects. Changing the size of units has a slightly larger effect but an order of magnitude lower than biases related to mis-specifications. Hence, choosing the right specification when measuring the impact of local characteristics appears to be more important than choosing the right spatial units. In practice, differences in estimates when the spatial scale varies can give a clue on the various agglomeration mechanisms at play at the various scales. Typically, knowledge spillovers, human capital externalities and matching effects should be the most prevalent agglomeration forces at short distances, say within cities or even neighbourhoods. By contrast, the effects of market access for both final and intermediate goods emphasised by economic geography models should be the main agglomeration forces driving differences in local outcomes at a larger scale, typically the region.

Keeping these remarks in mind, some articles have tried to evaluate the spatial extent of the impacts of local characteristics, and the scale at which they are the strongest. A common approach is to consider an individual or location defined at a fine scale and to draw rings around it with increasing radius. The value of any local characteristic can be computed using only locations within each ring separately. The spatial extent of agglomeration effects related to the local characteristic is then tested by including within the same specification its values for all rings. Among the first studies using this strategy on US data, Rosenthal and Strange (2003) were aiming at explaining local firm creation and Desmet and Fafchamps (2005) local employment. In Rosenthal and Strange (2003), local activity is considered to be located within 1 mile of the zip code centroid, and three rings around it are considered. The first ring contains activities located between 1 and 5 miles, the second between 5 and 10 miles, and the third between 10 and 15 miles. In Desmet and Fafchamps (2005), the first ring contains activities located between 0 and 5 kilometres from the county, the second between 5 and 10 kilometres, the third between 10 and 20 kilometres and so on every 10 kilometres up to 100 kilometres. Agglomeration effects are considered to attenuate with distance when a decreasing impact is obtained the further away the rings are from the location. The spatial scope of agglomeration effects is given by the distance after which the local characteristic does not have a significant effect any more. It can happen that agglomeration effects first increase with distance before decreasing. The turning point gives the spatial scale at which they are the strongest.

4.5.2 Measures of observed skills

Individual skills are not evenly distributed across locations. Combes et al. (2008a) show for instance that individual fixed effects and location fixed effects obtained from the estimation of a wage equation from French data are largely positively correlated. The uneven distribution of traits, intelligence and education is documented for the US by Bacolod et al. (2010). Bacolod et al. (2009b) show that city size is positively correlated with cognitive and people skills but negatively correlated with motor skills and physical strength. Bacolod et al. (2009a) also provide evidence that workers in the right tail of the people skill distribution in large cities have higher skills than those in small cities, and that the least skilled are less skilled in large cities than in small cities. This is in line with Combes et al. (2012c) who measure skills with individual fixed effects and Eeckhout et al. (2014) who measure skills with diplomas. Both papers conclude that there is a distribution of skills with larger variance and shifted to the right in larger cities. As discussed above, skills have two specific roles to play when estimating the effects of agglomeration economies on an economic outcome. First, skills can themselves be one of the determinants of agglomeration economies. Second, there can be some sorting of skills across locations and it is important to control for this to avoid biases when measuring the impact of local characteristics related to agglomeration economies.

As mentioned above, it is possible to keep the form of skills unspecified in wage equations by introducing individual fixed effects when using panel data. This has the two drawbacks that one has to rely on mobile individuals for identification and individual characteristics that matter for productivity cannot be identified. This strategy cannot be implemented when panel data are not available but various measures of observed skills can be used at the cost of not controlling for unobservable individual characteristics. There is a long tradition in labour economics of using obvious measures such as diplomas or years of schooling and we may mention Duranton and Monastiriotis (2002) on the UK and Wheaton and Lewis (2002) on the US as two early attempts that followed that route. It is also tempting to use the socio-professional category, 'occupation', which is often recorded in labour force surveys. It captures the exact job done by workers and part of the effects of past career, and may thus be considered as a measure that should be more correlated with current skills than education. On the other hand, there is an endogeneity concern since occupation is attached to the job and is jointly determined with the wage. There is no obvious solution for this endogeneity issue, except to use a more structural approach that would jointly model wages and occupational choice.

An interesting alternative is to introduce measures of traits and intelligence. Bacolod et al. (2009b) and Bacolod et al. (2010) build on psychological approaches and use detailed occupations from the Dictionary of Occupational Titles (DOT) to construct such measures using information on job requirements and principal component analysis. They end up with four indices related to cognitive skills, people skills, motor skills, and physical strength. It is possible to assess how individuals score on these four dimensions from the job they have just after their education. Bacolod et al. (2009b), in line with studies in labour economics, also use the Armed Forces Qualification Test (AFQT), the Rotter index, and the SAT scores for college admission in the US to control further for worker ability and better capture the quality of education. Some attempts have also been made to use other indirect proxies to control for skills. Fu and Ross (2013) use dummies for locations of residence, with the idea that the choice of a residential location is based on tastes, which are themselves likely to be partially correlated with individual productivity. At the same time, the location of residence can be endogenous as it is chosen while taking into account the location of the workplace and wage.

4.5.3 Functional form and decreasing returns to agglomeration

Most papers estimate a log-linear relationship between local outcome and local characteristics. When the elasticity is between 0 and 1, this corresponds to a function in levels which is concave but non-decreasing. This is an approximation and there is no theoretical reason why the relationship between the logarithm of local outcome and the logarithm of local determinants should be linear. Theory rather predicts that the marginal returns to agglomeration should be decreasing with city size, for instance because local congestion increases as the city grows. Gains from human capital

externalities from the first skilled workers in a location may be rather large, but the more numerous skilled workers are, the lower the marginal gain from one additional skilled worker. A similar line of argument may hold for most technological spillovers. Economic geography models with variable mark-ups and strategic interactions, such as the one proposed by Combes and Lafourcade (2011), do present the feature that in the short run gains from agglomeration dominate costs as long as the asymmetry between locations is not too large but further agglomeration in the largest locations can lead to a reverse result. As illustrated in Subsection 2.1, local productivity is negatively affected through some channels, such as the increase of land prices with population, whatever the city size. This kind of effect can become dominant when cities are very large. More generally, one expects gains from agglomeration to be increasing and concave with a steep slope at the beginning, and costs to be increasing and convex with an initial slope close to zero. In that case, the difference between the two is concave and bell-shaped. The relationship between the determinants of agglomeration economies, in particular population size, and local outcomes is then expected to decrease beyond some threshold.

The simplest way to test for the presence of non-log-linear relationships consists of augmenting the specification with the square of the logarithm of local determinants but more complex functions of local determinants such as higher order polynomials can also be used. For instance, Au and Henderson (2006b) regress the value added of a city on a non-linear specification of its size using a sample of Chinese cities. Graham (2007) develops an original strategy based on a translog production function and two measures of effective urban density. Effective density is computed as a market potential function using either straight-line distances or generalised transport costs that consider road traffic congestion. Corresponding measures are used to estimate the magnitude of diminishing returns from agglomeration, i.e. the concave impact of density, and its link with transport congestion. Note finally that the presence of concave effects can be studied for other local characteristics and outcomes. For instance, Martin et al. (2011) quantify the non-linear effect of specialisation on firm value added. Overall, the literature is rather suggestive of diminishing returns to agglomeration (see Section 5). In practice, when estimating a non-linear effect, one should always check that the support of observations covers the whole interval where the non-linear effect is interpreted. Otherwise, interpretation is based on extrapolation rather than an empirical feature of the data.

4.5.4 Spatial lag models

There is a strand in spatial econometrics considering that spatial lag models can be informative on the effect of local determinants of agglomeration economies. In these models, a local outcome is regressed on a weighted average of neighbours' outcomes or on a weighted average of neighbours' exogenous characteristics, or both, where weights decrease with distance, and the spatial correlation of residuals is sometimes taken into account (see Lesage and Pace, 2009, for details).

The weighted averages of neighbours' outcomes or characteristics are considered to capture agglomeration effects. It is now standard to estimate this kind of model with maximum likelihood. An important limitation to this approach is that the model is identified as a result of parametric assumptions, in particular as regards the impact of space on agglomeration effects and the distribution of residuals.

As emphasised by Gibbons and Overman (2012), spatial specifications face a reflection problem á la Manski, which is known to be very difficult to be properly dealt with. For instance, consider the case where individual wage is regressed on neighbours' composition in terms of diplomas because one expects human capital externalities to spill over the boundaries of spatial units. This composition may be endogenous as highly educated workers may be attracted to the vicinity of workers earning high wages, in particular because they can finance local public goods.

The reflection problem is usually addressed in spatial econometrics by using spatial lags of higher order as instruments, in the spirit of panel estimation strategies which consist in instrumenting variables by long time lags of their first difference. However, this kind of approach relies on assumptions on the extent of spatial effects. Indeed, one needs to assume that these effects only involve close neighbours, whereas more distant neighbours do not have any direct effect on the outcome, which is the reason why they can be used to construct instruments verifying the exclusion restriction. Nevertheless, it is possible that neighbours located further away also directly affect the outcome, and instruments are thus invalid. An additional issue is that the validity of instruments cannot be properly assessed using an over-identification test as all instruments are built from the same underlying variables, computed at various distances but fundamentally affected by common shocks.

Overall, the main identification concern remains: One needs to find a strategy to identify the effect of local determinants of agglomeration economies using a natural experiment or valid instruments, and unfortunately spatial lag models are of no help for that. Corrado and Fingleton (2012), Gibbons and Overman (2012), and McMillen (2012) propose a more thorough discussion of the concerns regarding spatial econometrics.

5 Magnitudes for the effects of local determinants of productivity

Previous sections present relevant strategies that could be used to estimate the impact of local determinants of agglomeration economies, and clarify the underlying econometric assumptions and interpretations. Contributions in the literature rarely adopt exactly these empirical strategies and often use variants. This makes it rather difficult to compare their results and it can sometimes explain discrepancies in their conclusions. We survey these contributions as well as their results, and try to emphasise the main assumptions that are made in the estimation strategies in light of previous sections. We first present the large body of articles on the average impact of density on

productivity. We then turn to the scarce papers estimating heterogeneous effects across city sizes, workers' skills, or industries. We also review contributions on the spatial extent of agglomeration effects, which include some using natural experiments to address endogeneity issues. Results on specialisation, diversity and human capital externalities are then exposed, and a final section is devoted to the results obtained for developing countries.

5.1 Economies of density

It is now established that the local density of economic activities increases the productivity of firms and workers. This conclusion emerges from a large number of studies quoted below. Some of them use aggregate data and regress the logarithm of regional wage or total factor productivity on current logarithm of employment or population density. Typical values for the elasticity when controlling for some local variables but ignoring both reverse causality and individual unobserved heterogeneity to deal with spatial sorting are between 0.04 and 0.07. Estimates are rather diverse because different countries, industries, or periods of time are considered, as emphasised by Melo et al. (2009). Some studies estimate even larger magnitudes but usually use fewer control variables. The elasticity range $0.04 - 0.07$ implies that when density is twice as great, productivity is between 3 and 5% higher. Density at the last decile in developed countries is usually at least two or three times greater than at the first decile, and may even be fifteen times greater (when considering European regions, or regions within some countries). The productivity gap associated to the inter-decile difference may be as large as 20%.

Correcting for aggregate endogeneity is generally found to have a small effect on elasticities. Instrumentation decreases them by 10 to 20%, and sometimes leaves the estimates unaffected or may even make them slightly increase. By contrast, using individual data and introducing individual fixed effects to control for spatial selection can change the estimated elasticity of productivity with respect to density much more. This elasticity can be divided by a factor larger than two and reach a value typically around 0.02. As detailed below, depending on the country and on the precise methodology used to control for skills (individual fixed effect or observed skills variables), the magnitude of the sorting bias can vary significantly.

Turning to specific estimates, the two benchmark studies using aggregate data for the US, Ciccone and Hall (1996) and Rosenthal and Strange (2008) for the years 1988 and 2000 respectively, report similar values for the elasticity of productivity with respect to density, at around 0.04-0.05. The first study instruments density with historical variables (for instance lagged population, lagged population density or lagged rail-road network) and the second study by geological variables (seismic and landslide hazard, percentage of area underlain by sedimentary rock). In both cases, instrumentation barely affects estimates, and if anything, slightly increases the elasticity of productivity with respect to density.

Some studies attempt to estimate this elasticity for European regions. Ciccone (2002) replicates

Ciccone and Hall (1996) on NUTS 3 regions in France, Germany, Italy, Spain and the UK. His main instrument is land area, which is not very convincing since we argue in Section 3.1 that land area can have a direct effect on productivity. He gets an elasticity of around 0.05 for 1992. Interestingly, he also finds no evidence that agglomeration effects significantly differ across countries. Two more recent studies extend the set of countries considered in the analysis, although at the cost of using larger spatial units. Brülhart and Mathys (2008) consider 245 NUTS 2 regions in 20 Western and Eastern European countries, with data on the 1980-2003 period for Western countries but only on the 1990-2003 period for Eastern countries, and 8 broad industries covering both manufacturing and financial services. They use first differences and GMM to deal with endogeneity issues in the estimations. Unfortunately, results seem to differ widely depending on the empirical strategy they adopt. Still, they estimate quite large agglomeration gains with a long-run elasticity of productivity with respect to density reaching 0.13. Interestingly, the strength of agglomeration effects seems to have increased over time. This result is consistent with economic geography models that predict a bell-shaped curve between trade costs and agglomeration gains. The European economy, which has experienced a decline in trade costs over the last decades, appears to lie on the right-hand side of the curve where agglomeration effects reinforce when trade costs get smaller. Foster and Stehrer (2009) obtain estimates closer to Ciccone (2002) when using a panel of over 255 NUTS 2 regions in 26 European countries for the 1998-2005 period that covers 6 industries, including "agriculture, forestry and fishing", which is not considered by Brülhart and Mathys (2008). They also obtain the further result of a larger magnitude of agglomeration economies for new member states than for old ones. Nevertheless, they use land area as the only exogenous instrument, as in Ciccone (2002), and consider that the regional skill composition is exogenous, which is not very convincing. Marrocu et al. (2013) further extend the number of countries, regions, and time-span while leaving aside the endogeneity issues and conclude that specialisation gains would be more prevalent in new member states and diversity in older ones.

A number of early studies estimate agglomeration economies for separate countries on either wages or TFP aggregated by region. We do not summarise the results of all these studies as they are already covered by Rosenthal and Strange (2004). We rather focus on recent articles that use richer datasets at the individual level that include workers' or firms' precise location.

Glaeser and Maré (2001) are the first to evaluate agglomeration effects on wages net of individual fixed effects, the analysis being conducted on US data. Unfortunately, the size of their dataset does not allow them to evaluate the elasticity of wages with respect to density but only the impact of a couple of dummies for city size. For the same reason, it is also difficult to compare the magnitude of the effects estimated by Wheeler (2006) and Yankow (2006), still from US data, to the rest of the literature. Combes et al. (2008a) are able to estimate the effect of density on wages across all French cities at the individual level while considering individual fixed effects and taking into account aggregate endogeneity with the two-step estimation procedure involving instrumentation

that is described in Subsection 2.1.1. They find an elasticity of wage with respect to density at around 0.030, which is half that obtained when individual unobserved heterogeneity is not taken into account. Using a more elaborate instrumentation strategy, Combes et al. (2010) obtain a value of 0.027. This figure is very close to the one obtained for Spain by De La Roca and Puga (2012) when they do not control for dynamic agglomeration effects, which is 0.025. Mion and Naticchioni (2009) replicate Combes et al. (2008a) with Italian data and get an even smaller estimate of 0.01 which is still significantly different from zero. From UK data, D'Costa and Overman (2014) get an elasticity of 0.016, and from Dutch data, Groot et al. (2014) get 0.021, controlling for many individual variables and city-industry-time fixed effects but not individual fixed effects.[15]

Combes et al. (2008a) also show that individual abilities do not distribute randomly across locations. Workers who have higher skills are more often located in the most productive cities, which are the densest. The correlation between individual and area fixed effects is 0.29, and the correlation between individual fixed effect and density is as high as 0.44. This is the fundamental reason why controlling for individual characteristics has so much influence on the estimate of the elasticity of productivity with respect to density. Mion and Naticchioni (2009) find that sorting is slightly weaker in Italy, as they obtain a correlation between individual fixed effect and density of 0.21. There is also some evidence of spatial sorting in Spain as shown by De La Roca and Puga (2012) when dynamic agglomeration effects are not taken into account, and in the UK as shown by D'Costa and Overman (2014) when both static and dynamic effects are considered.

The role of skills has been debated further by De La Roca and Puga (2012) who show from Spanish data that the explanatory power of individual fixed effects largely falls once dynamic agglomeration effects are taken into account in the specification. As detailed in Subsection 2.2, dynamic effects are captured with variables measuring the time spent in different classes of city size. When these variables are not included in the specification, having spent more time in larger cities is captured by the individual fixed effect. The inclusion of city experience variables allows the authors to disentangle the effects of individual skills captured by individual fixed effects from dynamic agglomeration gains. In order to assess the magnitude of dynamic gains, De La Roca and Puga (2012) consider a quantity defined at the city level as the sum of the time-invariant city fixed effect and the effect of experience accumulated in the city for a worker who stayed there for seven years (which is the average length of time for workers in their sample). The elasticity of this quantity with respect to density that captures both static and dynamic agglomeration effects is 0.049, which is almost twice as large as the elasticity of city fixed effects evaluated at 0.025. This indicates major dynamic gains which would be even larger for more able workers as shown by the estimation of a specification allowing for an interaction between the individual fixed effect and city experience. Perhaps surprisingly, dynamic gains are found to be independent of the size of the

[15] In contrast with these references, when considering individual data on siblings from the US, Krashinsky (2011) obtains that the average urban wage premium becomes non significant when introducing family fixed effects because there is a sorting of families across urban areas.

city to which workers move subsequently. There would thus be a homogeneous transferability of learning effects across locations.

Following an empirical strategy close to De La Roca and Puga (2012), D'Costa and Overman (2014) show for the UK that dynamic effects are also present but weaker than in Spain. In particular, dynamic gains appear to be one shot only, the first year of stay in a city, and do not cumulate over time (except for the youngest workers, below 21). These results are consistent with those of Faberman and Freedman (2013) who study the impact of firms' age on earnings returns to density with US data and find that almost all of the gains occur at firms' birth. The structural exercise conducted by Baum-Snow and Pavan (2012) allows them to consider endogenous individual location choices, static and dynamic heterogeneous agglomeration gains as well as matching effects. Their conclusions for the US are similar to those obtained for Spain. Both static and dynamic gains from agglomeration are present, static gains being more important to explain differences between small and medium cities, and dynamic gains playing a more significant role to explain differences between medium and large cities. Conversely, individual sorting and matching effects play a secondary role in the city wage premium.

Due to computation limits, many studies consider only classes of city size and not all the cities separately. Moreover, in De La Roca and Puga (2012), the heterogeneous individual impact of dynamic agglomeration economies is supposed to be identical to the direct effect of individual skills and static agglomeration effects are not allowed to be specific to skills, whereas in D'Costa and Overman (2014), both static and dynamic agglomeration effects are homogeneous across workers. Lastly, considering time-invariant city fixed effects makes the city experience component also capture the time evolution of static agglomeration gains. Other recent attempts that consider both static and dynamic effects in specifications closer to those of Glaeser and Maré (2001) include Lehmer and Möller (2010) who find for Germany that only dynamic effects occur once firm size and individual fixed effects are taken into account, Carlsen et al. (2013) who find for Norway that static gains are homogeneous across education levels while dynamic ones increase with education, Wang (2013) who finds for the US that both static and dynamic gains are present and that they are stronger for younger and more educated workers. To conclude, De La Roca and Puga (2012) and Baum-Snow and Pavan (2012) pioneered the simultaneous study of static and dynamic agglomeration effects on wages while taking into account the observed and unobserved heterogeneity of workers. Further investigation along the lines suggested in Section 2 constitutes an appealing avenue of research.

As discussed in Subsection 4.1, it is worth studying TFP rather than wage since it is a direct measure of productivity that can sometimes be computed at the firm or establishment level, keeping in mind that interpretations slightly change. On the other hand, no convincing method has been proposed to control for individual skills when estimating agglomeration effects on TFP even with individual data at hand, and we have seen that sorting according to skills can induce considerable

biases. Henderson (2003) for the US and Cingano and Schivardi (2004) for Italy are among the first papers to study firm-level TFP. However, their assessment of possible endogeneity biases is only partial. Henderson (2003) uses GMM techniques to instrument both input use and local variables, with the caveats we mentioned in Subsection 4.3.3. Cingano and Schivardi (2004) takes into account the endogeneity of input use only, through the use of the Olley-Pakes estimation procedure. Graham (2009) provides estimates for the UK based on firm-level TFP data but he instruments neither input use nor local effects. Di Giacinto et al. (2014) assess the respective impact of locating in an urban area and in an industrial district on firm-level TFP in Italy, while instrumenting input use but not the size of the local economy that is also included as a control. As regards France, Combes et al. (2010) estimate firm TFP with the Olley-Pakes estimation procedure among others and use the estimates to construct a local measure of TFP which is then regressed on density while using historical and geological variables as instruments. Martin et al. (2011) rather rely on GMM using lagged values of explanatory variables as instruments. To the best of our knowledge, a large number of European countries, including Germany and Spain, have not yet benefited from specific estimates of agglomeration effects on TFP.

Studies on TFP usually conclude that there are significant agglomeration gains on firm productivity, even if some authors that simultaneously control for the level of industrial employment (not its share) wrongly reach the conclusion of their absence (see the discussion in Subsection 3.2). Melo et al. (2009) show that elasticities of TFP with respect to density are on average estimated to be larger than those obtained for wages, typically around 50% larger, even in Combes et al. (2010) where both types of estimates are computed on the same dataset and endogeneity is taken into account using the same instruments. Indeed, these authors get an elasticity of TFP with respect to density of 0.035-0.040 whereas they obtain 0.027 for the elasticity of wages. According to our basic model, it is difficult to interpret the difference between the two types of estimates. In wage equations, all the effects are re-scaled by the share of labour in the production function. Moreover, agglomeration economies percolating through the cost of inputs other than labour, such as land and intermediate inputs, affect wages but not TFP (see Subsection 4.1). A further possible reason for the difference in estimates obtained from wage and TFP regressions is that no one has managed to successfully control for individual skills when working on TFP. Taking properly into account workers' unobserved heterogeneity in TFP estimations is an avenue for future research.

5.2 Heterogeneous effects

As explained in Subsection 4.5.3, the impact of local characteristics on productivity should be bell-shaped as agglomeration gains are increasing and concave, while agglomeration costs are increasing but convex. Variations in the marginal effects of local characteristics are a first type of heterogeneity. For instance, the gain from increasing city size could be positive and large for small cities, and turn negative for very large ones, predictions that need to be investigated for instance

to assess whether or not the size of cities is optimal.

Most studies do not report an estimated degree of concavity for agglomeration effects. Exceptions include Au and Henderson (2006b) who estimate for China a bell-shaped relationship between the productivity and size of cities and conclude that most cities lie on the left-hand-side of the peak, i.e. are too small to achieve the highest level of productivity. For the UK, Graham (2007) develops an original strategy based on road traffic congestion to estimate the diminishing returns of agglomeration effects and their link with transport congestion. Five out of nine industries present concave effects of density. Furthermore, it is shown that when congestion is taken into account, the elasticity with respect to density increases in seven of the nine industries. This is in line with expectations since in the absence of controls, the elasticity with respect to density reflects the overall net impact of density, taking into account both positive and negative effects. In the UK, congestion is shown to represent up to 30% of the agglomeration effect.

Agglomeration effects can also be heterogeneous across industries as the strength of agglomeration economies depends on industry characteristics. Nevertheless, estimations by industry remain scarce. One reason may be that the design of the empirical model, and in particular the search for valid instruments, has to be done industry by industry. Another reason is the lack of availability of local data per industry. Brülhart and Mathys (2008) and Foster and Stehrer (2009) are notable exceptions, but their works are at the European regional level and do not control for individual effects. They find significant agglomeration effects in all but one of the industries they consider. The exception is agriculture, in which regional density has a negative impact, a result that is fairly intuitive. Given the share of land in agricultural production and the fact that land prices increase with density, less dense places clearly represent the best alternative for productivity in this industry. Morikawa (2011) estimates from firm-level data the elasticity of firm TFP with respect to density for detailed services industries in the US without instrumenting. He finds large elasticities ranging from 0.07 to 0.15. In their meta-analysis, Melo et al. (2009) conclude that on average agglomeration effects tend to be stronger in manufacturing industries than in services industries.

Some studies have tried to evaluate the extent to which agglomeration economies are stronger for some types of workers or firms. For instance, Bacolod et al. (2009a) and Abel et al. (2012) for the US, Di Addario and Patacchini (2008) for Italy, and Groot and de Groot (2014) for the Netherlands, confirm the intuition that returns to education are higher in cities. This is also found by Lindley and Machin (2014) for the US who then assess to what extent the change in wage inequality across States over the 1980-2010 period arises from a shift in skill composition and a variation in education-specific returns to agglomeration economies. Firms in industries that are more skill-intensive should be concentrated where returns to education are higher, the larger cities, and this is observed by Elvery (2010) for US metropolitan areas.Lee (2010) is one of the rare studies to exhibit an industry in which the urban wage premium is found to decrease with skills, the health care sector in the US. He explains his result by labour supply effects for high-skilled

health care employees as surgeons, dentists, or podiatrists, who would be more attracted by urban life than nurses or massage therapists, and this woud put a downward pressure on their wages in larger cities.

Using a structural approach controlling for endogenous location choices, Gould (2007) shows that both static and dynamic agglomeration gains are present for white-collar but not for blue-collar workers. Matano and Naticchioni (2012) reach a similar conclusion after performing quantile regressions on Italian data controlling for sorting on unobservable worker characteristics. They find that agglomeration effects appear to strengthen along the wage distribution. This is in line with the conclusions of Combes et al. (2012b) who use the full distribution of firm level TFP in France to show that the most efficient firms gain more from density than the least efficient ones. For instance, firms in the last quartile of productivity gain three times more from density than those in the first quartile. It is also found that the largest establishments gain more from density. Benefits are 50% greater for establishments with more than 100 workers than those with 6 to 10 workers. Going in the opposite direction, Henderson (2003) and Martin et al. (2011) conclude that specialisation effects are larger for smaller firms, but these two papers measure specialisation with the level and not share of industrial employment. Therefore, they partially confound density and the specialisation effects as explained in Subsection 3.2.

Other authors have investigated the sources of heterogeneous productivity gains from agglomeration, but rarely take into account simultaneously the endogeneity issues related to reverse causality and missing local variables. For instance, Rosenthal and Strange (2003) using US data find that the number of hours worked decreases with density for non-professionals but increases for professionals, and the effect is stronger for young workers. Moreover, the number of hours worked by young professionals is particularly sensitive to the proximity of other young professionals. Bacolod et al. (2009b) investigate which skills have returns positively related to city size. They conclude that only cognitive and social skills are better rewarded in large cities, while motor skills and physical strength are rewarded less well. In line with these results, Andersson et al. (2014) find that it is only for non routine jobs that there are gains from agglomeration in Sweden once the spatial sorting of skills is taken into account.

There is also scarce evidence on heterogenous agglomeration gains across demographic groups. Phimister (2005) estimates gender differences in city size premium from UK data, controlling for individual fixed effects but without taking into account endogeneity issues. He finds a larger urban premium for women, especially for those married or cohabitating. Ananat et al. (2013) investigates differences across races in the US while controlling for unobserved worker heterogeneity through residential location choices as in Fu and Ross (2013) but without dealing with endogeneity issues at the local level. They find that agglomeration effects are heterogeneous across races, the black-white wage gap increasing by 2.5% when there are one million more inhabitants in the city.

5.3 Spatial extent of density effects

The rapid spatial decay of agglomeration effects is another robust finding in the literature. Agglomeration economies do not spill much over space. For the advertising agency industry, Arzaghi and Henderson (2008) provide evidence of an extremely fast spatial decay of agglomeration effects that are shown to occur primarily within 500 metres. This decay is certainly too extreme to be representative of more standard industries but, still, effects are rarely found to be significant beyond 100 kilometres, and the threshold is often lower.

The first way to assess the spatial extent of agglomeration effects consists of considering a single market potential variable that encompasses both the own location size and the sizes of other locations. As detailed in Subsection 3.1, one can consider the Harris market potential which is simply the sum over all spatial units, including the own location, of their size (or density) divided by the distance between the location and the unit considered. More structural forms of market potential from economic geography models can also be used. Importantly, in all cases, one implicitly assumes a pretty strong spatial decay of agglomeration effects. For instance, when trade costs are inversely related to distance, the impact on a location of the economic activity located 20 kilometres away is four times lower than that of activity at 5 kilometres, it is 10 times lower at 100 kilometres than at 10 kilometres, and so on. The positive effect of the economic size of distant locations, and the spatial decay of this effect are rarely rejected empirically. For instance, Head and Mayer (2006) in a study on European NUTS 2 regions obtain, when neither local skills nor endogeneity are taken into account, that both the Harris market potential and a structural market potential significantly increase regional wages, the two variables having a similar explanatory power. Holl (2012) assesses the effect of a Harris market potential based on distance through the real road network which is instrumented with historical population, geology, and historical transport networks. He finds a positive effect of this market potential on regional wages in Spain. Structural papers following the one proposed by Hanson (2005), such as the two early replications by Mion (2004) for Italy and Brakman et al. (2004) for Germany, confirm the positive impact of structural market potential on regional wages, even if sorting on skills and endogeneity concerns are not always fully addressed. Brakman et al. (2006), Breinlich (2006), Brakman et al. (2009) and Bosker et al. (2010) find evidence of a positive effect of structural market potential on GDP per capita for NUTS 2 European regions. Fallah et al. (2011) show for US metropolitan areas that the impact of the structural market potential is stronger at the top of the wage distribution. Some other contributions for developing countries are discussed in Subsection 5.7.

Assessing separately the role of own density and market potential definitely makes more sense if different local externalities operate at different distances. External market potential (that excludes own size or density) is most often found to have a significant positive effect on local productivity when it is introduced in addition to density in the specification. For instance, Combes et al. (2008a) and Combes et al. (2010) find that both variables have a significant positive effect in France, even

when they are both instrumented and individual unobserved heterogeneity is taken into account. For NUTS 2 European regions, Foster and Stehrer (2009) introduce next to density a measure of market potential with a spatial decay of agglomeration economies arising from other regions of exponential form, i.e. with a decline that is even sharper than the inverse of distance. When trying exponential functions with various coefficients, they find that only those with the strongest spatial decay exhibit significant effects. Note that, in general, introducing the external market potential in regressions only slightly reduces the impact of own density.

The second strategy for assessing the spatial decay of agglomeration economies consists of introducing in the specification variables for the economic size of distant locations. Ciccone (2002) finds for NUTS 3 European regions that production in neighbouring regions has a positive impact on local productivity. He does not report the magnitude of the coefficient however, and he does not test for an impact of regions located further away. Rice et al. (2006) find for UK regions that agglomeration economies attenuate sharply with distance. Distant markets do affect local wages and productivity, but markets located 40 to 80 minutes away have one-quarter the effect of those located at less than 40 minutes, and markets located 80 to 120 minutes away have half the effect of those located 40 to 80 minutes away. Moreover, there is no effect of markets located more than 80 minutes away. Rosenthal and Strange (2008) obtain even larger spatial gradients when estimating the effect of employment concentration in rings around location on wages in US cities. The effect of the 0-5-mile ring is four to five times larger than the effect of the 5-25-mile ring. Turning to the outer rings (25-50 miles and 50-100 miles), effects are even smaller and very often not significantly different from zero. The spatial pattern obtained for Italy by Di Addario and Patacchini (2008) is consistent with this one since the impact of local population size is strongest between 0 and 4 kilometres and is not significant any more beyond 12 kilometres.

5.4 Market access effect evaluated using natural experiments

As our chapter shows, strategies used to tackle endogeneity issues are not always convincing, and in some cases, authors do not even attempt to tackle them. A few recent publications propose using natural experiments as a source of variation in the local economy size to circumvent endogeneity problems. Greenstone et al. (2010) test the presence of agglomeration effects on firm TFP by exploiting the arrival of large plants in some given US counties. Such plants affect the intensity of agglomeration economies although it is not possible to quantitatively assess the exact magnitude of the shocks. The key idea for finding a relevant control group for counties receiving a large plant is to rely on a real estate journal, the *Million Dollar Plants*, that gives for any large plant created, the county that the plant ultimately chose (the winner) and the counties that survived a long selection process but were ultimately not selected (the runners-up). The authors show that on average runner-up counties have similar characteristics to the winners. The effect of plant arrivals on incumbent plants is studied in a panel including both winner and runner-up counties but not

others. Firm TFP is regressed on an interaction term between a dummy for being in the winner group and a dummy for the dates after the arrival of the large plant. The estimated coefficient of this interaction corresponds to the difference-in-differences estimator. It is found to be significantly positive and sizable, especially for incumbent plants sharing similar labour and technology pools with the new plant. Whereas the empirical strategy is quite convincing for identifying the effect of arriving plants, the link between the arrival of plants and changes in the intensity of agglomeration spillovers remains unknown (see the argument in Subsection 4.3.4). Moreover, external validity is far from granted since only a small sub-sample of counties is studied.

Papers exploiting natural experiments to evaluate the effect of market potential typically use the opening and closing of frontiers that prevent firms or cities from interacting with neighbours. An early example is Hanson (1997) who studies the effect of the trade reform in Mexico in the 1980s that turned the country from a closed economy to an economy open to trade with foreign countries, and in particular with the US. The opening of the frontiers has increased the market potential, especially for firms close to the Mexican-US border. It is shown that the opening of frontiers attracted firms close to this border whereas the concentration of firms in the capital city Mexico, which is located at a distance from this border, decreased. A more recent interesting use of natural experiment is provided by Redding and Sturm (2008) who study the effect of the division of Germany in 1949 on the growth of cities on the western side of the West-East German border.[16] The border cut their access to cities on the eastern side and thus decreased their market potential. The effect on cities located further away from the border should have been smaller as they had better access to other cities in Western Europe. Consequently, the authors compare the population growth of western cities close to the border to that of western cities far from the border, the two groups of cities having the same population trends before the division of the country. This is done in the same spirit as Greenstone et al. (2010), by restricting the sample to western cities and regressing city growth on an interaction term between a dummy for being close to the West-East German border and a dummy for dates after 1949. It is found that division of Germany led to a substantial relative decline of population growth for cities close to the border.[17] The effect is larger for smaller cities, which is expected since they have a smaller own market and rely more on other city markets. An interesting additional exercise would be to assess to what extent the division of Germany decreased the value of a market potential index and deduce from this measure of the shock and the difference-in-differences estimator a value for the elasticity of population growth with respect to market potential. This coefficient could be compared to the one obtained using a

[16] Note that the outcome here is city growth and not productivity as in other contributions surveyed in this section. This is because we chose to review all significant papers using natural experiments at the same place. Other results on city growth are reviewed in Section 6.

[17] A follow-up study (Ahlfeldt et al., 2012) shows that the division and reunification of Berlin had a significant effect on the gradient of land prices and employment in West Berlin close to the former main concentration of economic activity in East Berlin but a negligible effect along other more economically remote sections of the Berlin Wall.

more standard least squares instrumentation approach.

5.5 Specialisation and diversity

We now review papers evaluating the effect of localisation economies on local productivity. The main variable used for that purpose is specialisation, which is computed as the share of the industry in the local economy. Its effect on local productivity is assessed while controlling for the size or density of total activity. In many studies, when density and specialisation are simultaneously introduced, both are found to have a significant positive effect on productivity. For instance, Cingano and Schivardi (2004) show that this is the case in Italy when industries are pooled together. They also find that the spatial decay is very strong, since specialisation in neighbouring regions has no impact on local productivity. For France, Combes et al. (2008a) find that the effect of specialisation, estimated on wages separately for each industry, is significantly positive for 94 industries out of 99. Its magnitude is larger in business services and in two high-tech industries, medical instruments and artificial fibres. This is intuitive since such industries could face stronger technological spillover effects. These results confirm those of Henderson (2003) for the US where a larger effect of specialisation is found in high-tech industries. Martin et al. (2011) obtain a significant positive effect of specialisation on firm productivity in France that becomes negative above a certain level of specialisation, which is consistent with the presence of concave localisation effects. From European data, Brülhart and Mathys (2008) find a negative impact of own-industry density on output per worker in the industries they study, to the notable exception of financial services. Using a spatial variance analysis, Combes et al. (2008a) show that whereas total employment density explains a large share of spatial disparities in productivity, the explanatory power of specialisation remains small.

Following both the intuition of Jacobs (1969) and the central role of preference for diversity in many economic geography models, another appealing variable to explain productivity is the overall industrial diversity of the location. However, its estimated effect has been shown to be not robust. It is sometimes significantly positive, sometimes significantly negative, and often not significant at all, as for example for France in both Combes et al. (2008a) and Combes et al. (2010), for Italy in Cingano and Schivardi (2004) and for the US in Henderson (2003). Even if there are interesting intuitions behind diversity variables, no effect seems to be at play. It may be due to the way diversity is measured, since it is often through a Herfindahl or Krugman specialisation index computed from the industry shares in the local economy using a rather aggregate industry classification. Moreover, some industries may benefit from a group of other industries but usually not from all industries as assumed in the Herfindahl index. To tackle this issue, Moretti (2004a) uses a measure of proximity between industries and finds for the US that spillovers between economically-close industries are larger than spillovers between economically-distant industries, and this matches better what Jacobs had in mind.

5.6 Human capital externalities

We have already emphasised that the local share of professionals or highly-educated workers has many effects on productivity that can be difficult to disentangle. First, when using data aggregated at the city or the region level, it is not possible to identify separately the direct composition effect of skilled workers on average productivity and their human capital externality effect. When using individual data, the role of the local share of skilled workers on individual productivity can be assessed, while simultaneously taking into account the direct composition effect by introducing individual variables or individual fixed effects. Nevertheless, Section 3.3 shows that the local share of skilled workers captures not only the externality effect but also a substitution effect which is positive for unskilled workers and negative for skilled workers.

There has been a debate since the beginning of the 2000s on the existence and magnitude of local human capital externalities. While Moretti (2004a) and Moretti (2004b) find significant positive effects of human capital measures, Ciccone and Peri (2006) rather obtain an estimate that is not significant. It is difficult to make a conclusive case for either side. Moretti (2004b) implements the now standard approach of regressing the individual wage on the share of college-educated workers but this share captures both the externality and substitution effects. This is also the case in Moretti (2004a) when studying TFP rather than wages. On the other hand, Ciccone and Peri (2006) use a shift-share approach supposed to control for substitution effects but the sources of identification remain unclear as explained in Subsection 3.3. Importantly, no paper simultaneously controls for the presence of possible gains from density, whereas density is usually positively correlated with local human capital.

Other papers mostly use the same approach as Moretti (2004b) and obtain similar results. Rosenthal and Strange (2008) find the same positive effect of the local share of college-educated workers in the US. Considering this share at various distances from each worker location, they also obtain that the effects of human capital externalities attenuate sharply with distance. The effect of the share of college-educated workers in the 0-5-mile ring around the location is 3.5 times larger than the effect of this share in the 5-25-mile ring. These results are consistent with those of Fu (2007) who finds for the Boston Metropolitan Area using data on census blocks that human capital externalities decrease quickly beyond three miles.

For Europe, Rice et al. (2006) assess the role of the local share of workers with degree-level qualifications in the UK and find that it has a positive effect on wages and productivity. However, since the specification is not estimated at the individual level but rather at the local level, it is not possible to quantify separately the composition and externality effects. This is possible for France and Combes et al. (2008a) find a positive effect of the local share of professionals within the industry on individual wages, even after controlling for individual fixed effects and age, as well as location-time fixed effects which capture in particular the effect of density. Similarly, Rodríguez-Pose and Tselios (2012) find a positive impact of the regional levels of education on

individual earnings for European regions while using individual data and controling for individual characteristics and a region-time fixed effect.

Interestingly, when both productivity and wage data are available, one can evaluate how much of the productivity gains due to agglomeration are transformed into wage gains for workers. While this has not been done for Europe, Moretti (2004a) finds for the US that estimated productivity differences between cities with high human capital and low human capital are similar to observed differences in wages of manufacturing workers, indicating an almost complete transfer of human capital effects to workers. Since unobserved worker heterogeneity is not controlled for in this paper, the similarity between the productivity and wage differences can also result from a composition effect affecting both wage and TFP.

5.7 Developing economies

We now present empirical results on the presence of agglomeration economies in some developing countries. The related literature is recent and research needs to be pursued to gain knowledge on additional countries. The effect of market size on wages has been studied for China, India and Colombia. Panel data are usually not available and it is thus in general not possible to take into account unobserved individual heterogeneity. Differences between individuals are rather taken into account through individual explanatory variables such as qualification, gender, age, and sometimes occupation or the type of firm where the individual is employed. Overall, market size is found to have a larger effect than in developed countries. Combes et al. (2013) for instance study the effect of density on individual wages in 87 Chinese prefecture cities, instrumenting density by peripherality, the historical status of the city, and distance to historical cities. The elasticity of wages with respect to density is found to be at 0.10-0.12, around three times larger than in developed countries. Chauvin et al. (2014) evaluate the effect of density on individual annual earnings in India at the district level and also find a large elasticity around 0.09-0.12. Duranton (2014) investigates the impact of population on individual wages in Colombia while controlling for area at the local labour market level (which amounts to investigating the effect of density). Instrumentation is conducted using historical populations or soil characteristics (erodability and fertility). The estimated elasticity is 0.05, and thus lower than in China and India, but still large compared to estimates for developed countries.

Other measures of productivity have been used. Henderson et al. (2001) evaluate the effect of city population on value added per worker in Korea for five industry groups and 50 cities using panel data over the 1983-1993 period. They do not find evidence of size effect for any industry but their results are based on time evolutions without instrumentation for the endogeneity of city population. Similarly, Lee et al. (2010) find that population density does not have any significant effect on establishment-level output per worker in Korea when estimating a specification where local fixed effects and control variables are considered. Au and Henderson (2006a) and Au and

Henderson (2006b) study at the city level the effect of total employment and its square on output per worker in China in the 1990s, instrumenting with urban plans not related to output and urban amenity variables. They control for the local shares of manufacturing and services and the shape of the total employment effect is allowed to vary with these shares. They find a concave effect of total employment on output per worker. The vast majority of Chinese cities appear to have a size of less than 50% of the peak where agglomeration economies are the most important. This can be explained by the *hukou* system that restricts workers' social rights mostly to their birthplace and thus limits their mobility, especially in the 1990s when it was strictly enforced.

There are also a couple of publications on firm productivity. Lall et al. (2004) study the effect of urban density on firm productivity in India for eleven industries considered separately, estimating jointly a production function and a cost function. The effect is found to be significantly positive in one industry only. Saito and Gopinath (2009) quantify the impact of regional population on firm TFP in the food industry in Chile estimating a production function using the Levinsohn-Petrin approach. The elasticity is found to be significantly positive, at around 0.07. In both papers, the authors do not deal with the endogeneity of local determinants of agglomeration economies.

The role of market potential is considered along with the size of the local economy by some of the previous articles. Lall et al. (2004) study the impact of the Harris market potential in India, an originality of their work being the use of accurate transport times rather than distances in the construction of their market potential variable. This variable includes the own location and its effect is found to be negative but non-significant for several industries. Other papers conduct similar exercises but removing the own area from the computation of the market potential measure to disentangle the size effects from the local economy and external markets. Interestingly, Duranton (2014) obtains a significantly negative sign for the effect of external market potential on wages in Colombia. An explanation can be that, when workers are perfectly mobile as in Krugman (1991a), the spatial equilibrium without full agglomeration implies lower nominal wages in larger regions to compensate for the better market access that decreases the prices of consumption goods. Combes et al. (2013) find no significant effect of market potential on wages in China once it is instrumented simultaneously with other local determinants, whereas Au and Henderson (2006a) find a positive effect on output per worker but the variable is not instrumented.

Some papers have adopted quasi-structural approaches inspired by Redding and Venables (2004) and Hanson (2005) to focus on the effects on wages of structural market access and supplier access that are derived from economic geography models. It has the limitation that own area is involved in the construction of the access variables and the effect of own local economy size cannot be identified separately from the effects of external market and supplier access. Amiti and Cameron (2007) study the effect of both access variables on wages at the firm level in Indonesia, but without being fully structural in their construction and without instrumenting to take into account endogeneity issues. Both market and supplier access are found to have a positive effect.

Only 10% of the market access effect goes above 100km, and only 10% of the supplier access effect goes above 262km.

Fally et al. (2010) evaluate the impact of market and supplier access on individual wages in Brazil using a two-stage approach. First, a wage equation including state-industry fixed effects and individual characteristics is estimated in the spirit of Combes et al. (2008a) but at the industry level and without individual fixed effects since only cross-section data are available. In a second step, estimated state-industry fixed effects are regressed on structural measures of market and supplier access. These measures are obtained following strictly the strategy proposed by Redding and Venables (2004) where market and supplier access are recovered from the estimates of the trade flow specification derived from a economic geography model. An originality is that trade flows are measured at the industry level, which allows the authors to construct the access variables for each industry separately, whereas other papers only use aggregate flows and therefore construct only aggregate access variables.[18] Both market and supplier access variables are found to have a significant positive effect on wages when estimations are conducted using OLS. The authors then remove the supplier access variable from the specification and instrument the market access variable only (both variables rarely have simultaneously a significant effect due to their high correlation). Market access is found to keep its significant positive impact on wages.

Finally, Hering and Poncet (2010) evaluate the effect of market access on individual wages in 56 Chinese cities. They also follow the strategy proposed by Redding and Venables (2004) to build the market access variable but they do not consider the role of supplier access at all. Labour skills are captured by individual observed characteristics and a single-step estimation strategy is used. The authors instrument market access by centrality indices and find a significant positive effect which is larger for skilled workers.

Note that in all these contributions, structural access variables are the only local determinants of agglomeration economies considered in the specifications. Therefore their impacts cannot be identified separately from the effects of other local determinants not derived from economic geography models if these other determinants are correlated with access variables, which can occur in particular when distance plays a similar role in the attenuation of their effects.

Finally, some papers have studied local determinants of agglomeration economies other than market size. Henderson et al. (2001) assess the effect of industrial specialisation (measured with industry local employment) on productivity growth in Korea. They find some evidence of localisation economies for all the industry groups they consider, the magnitude of the effects being similar to that of the US. Lopez and Suedekum (2009) are interested in localisation economies and agglomeration spillovers on TFP for establishments in Chile. They consider both downstream and upstream spillovers between firms related by input-output relationships. They find a positive

[18]The authors could have conducted the second step estimation for each industry separately, as proposed in Subsection 2.1, but they prefer to pool all industries together possibly because the number of locations (27 states) is small.

effect of the number of intra-industry establishments consistent with the presence of localisation effects and of the number of establishments in upstream industries consistent with unidirectional agglomeration spillovers. Saito and Gopinath (2009) evaluate the impact of diversity, measured by a Herfindahl index, on firm TFP in the food industry in Chile, but find no significant effect. Endogeneity of local determinants and spatial sorting of workers are considered in none of these papers.

6 Effects of agglomeration economies on outcomes other than productivity

Although the most straightforward interpretations are made for the effects of local variables on local productivity, a rather large literature has attempted to identify the role of agglomeration economies on local outputs other than productivity. These outputs include employment or employment growth, and firm location decisions. We now turn to this literature and relate it to the same theoretical framework as the one we developed for productivity. This allows us to emphasise difficulties that are encountered when interpreting the results. Nevertheless, we survey the results that have been obtained over the last decade.

6.1 Industrial employment

We first focus on the local determinants of local industrial employment. We provide a theoretical background to specifications estimated in the literature, comment on the interpretations that can be made of the estimated coefficients, and finally present the results obtained by related papers.

6.1.1 From productivity externalities to employment growth

The two early studies that initiated the empirical evaluation of agglomeration economies in the 1990s, Glaeser et al. (1992) and Henderson et al. (1995), do not directly focus on the determinants of local productivity but rather on those of local employment growth at the industry level. A possible reason is that data on wages or total factor productivity at fine geographical levels such as cities or local labour markets were less available than today, and this is even more the case for individual data. At the same time, employment is by itself a local outcome of interest, especially for policy-makers, when for instance regional unemployment disparities are large as in Europe.

We develop a theoretical framework similar to the one used for productivity in order to ground employment equations and to allow for relevant interpretations of the effects found in this literature. As will become clear below, it is necessary to rely on a production function at the industry level

with non-constant returns to scale and we consider:

$$Y_{c,s,t} = \frac{A_{c,s,t}}{\alpha_1^{1-\alpha_2}\alpha_2^{\alpha_2}} \left(s_{c,t}L_{c,s,t}\right)^{\alpha_1} K_{c,s,t}^{\alpha_2}, \tag{57}$$

where $\alpha_1 + \alpha_2 < 1$. The first-order conditions equalising the return of inputs to their marginal productivity are:

$$w_{c,s,t} = \frac{\alpha_1 p_{c,s,t} A_{c,s,t}}{\alpha_1^{1-\alpha_2}\alpha_2^{\alpha_2}} s_{c,s,t}^{\alpha_1} L_{c,s,t}^{\alpha_1-1} K_{c,s,t}^{\alpha_2}, \tag{58}$$

$$r_{c,t} = \frac{\alpha_2 p_{c,s,t} A_{c,s,t}}{\alpha_1^{1-\alpha_2}\alpha_2^{\alpha_2}} s_{c,s,t}^{\alpha_1} L_{c,s,t}^{\alpha_1} K_{c,s,t}^{\alpha_2-1}. \tag{59}$$

Substituting into (59) the expression of capital given by (58) leads to:

$$L_{c,s,t} = \left(\frac{p_{c,s,t} A_{c,s,t} s_{c,s,t}^{\alpha_1}}{w_{c,s,t}^{1-\alpha_2} r_{c,s,t}^{\alpha_2}}\right)^{\frac{1}{1-\alpha_1-\alpha_2}} \tag{60}$$

We first leave aside the role of wages that will be discussed below. Making the same assumptions as in Section 2 on how local characteristics determine $p_{c,s,t}$, $A_{c,s,t}$ and $r_{c,s,t}$, equation (60) can be used to motivate an empirical specification where the logarithm of local industry employment (instead of wage) is expressed as a function of local variables such as local density, land area, and specialisation:

$$\ln L_{c,s,t} = \beta \ln den_{c,t} + \mu \ln area_{c,t} + \vartheta \ln spe_{c,s,t} + \nu_{c,s,t}\,. \tag{61}$$

First notice that, as in the case of productivity, the exact channel of agglomeration economies cannot be identified since local characteristics determining agglomeration effects may have an impact on employment not only through technological progress, but also through input prices and goods prices. Importantly, the role of specialisation cannot be identified since the dependent variable, industrial employment, is a log-linear combination of specialisation and density, and terms have to be rearranged to avoid redundancy.This identification issue is the reason why the production function was specified at the industry level. By contrast, the role of other local variables can still be studied since (61) implies:

$$\ln L_{c,s,t} = \frac{\beta - \vartheta}{1-\vartheta} \ln den_{c,t} + \frac{\mu - \vartheta}{1-\vartheta} \ln area_{c,t} + \nu_{c,s,t}\,. \tag{62}$$

The impact of remaining local determinants is now net of the impact of specialisation, and cannot be identified separately from it.[19] It was initially suggested in the literature that the static agglom-

[19]Firm level data would make it possible to identify the effect of industry employment by regressing firm employment on industry employment, in a way analogous to how individual wages allowed us to identify the role of individual skills separately from human capital externalities. This has not been done before to the best of our knowledge.

eration effect related to specialisation could be identified using non-linearities by also including in (62) the level of specialisation in addition to its logarithm as an extra local variable. However, this makes interpretations difficult, especially when the two effects are estimated with different signs as for instance in Henderson et al. (1995). Parametric identification relying only on specific functional forms should be avoided.

Glaeser et al. (1992) propose rewriting (60) in first difference and then considering that the growth rate of local variables instead of their level is a function of the levels of local determinants. They interpret local variables as determinants of technological progress but these variables also capture the role of agglomeration economies operating through goods and inputs prices as shown by (60). Specialisation can now be included among local characteristics and its effect is identified separately. The corresponding specification is given by:

$$\ln L_{c,s,t} - \ln L_{c,s,t-1} = \widetilde{\beta} \ln den_{c,t-1} + \widetilde{\mu} \ln area_{c,t-1} + \widetilde{\vartheta} \ln spe_{c,s,t-1} + \varepsilon_{c,s,t} \,. \tag{63}$$

The coefficients of local variables capture the impact of dynamic agglomeration effects such as improved learning but not the impact of static ones as in (62).

When there is time autocorrelation of residuals, it is possible to derive from (62) a dynamic specification of local-industry employment similar to (63) even if there are no static and dynamic agglomeration effects. Suppose for instance that $\nu_{c,s,t}$ follows an AR(1) process such that:

$$\nu_{c,s,t} = (1-\rho)\, \nu_{c,s,t-1} + \varepsilon_{c,s,t}, \tag{64}$$

where $0 < \rho < 1$ and the residuals $\varepsilon_{c,s,t}$ are identically and independently distributed. When there is no agglomeration effect such that $\nu_{c,s,t} = L_{c,s,t}$ and taking into account the fact that $L_{c,s,t} = den_{c,t}\ area_{c,t}\ spe_{c,s,t}$, this specification implies:

$$\ln L_{c,s,t} - \ln L_{c,s,t-1} = -\rho \ln L_{c,s,t-1} + \varepsilon_{c,s,t} = -\rho \ln den_{c,t-1} - \rho \ln area_{c,t-1} - \rho \ln spe_{c,s,t-1} + \varepsilon_{c,s,t} \,, \tag{65}$$

which involves the same explanatory variables as (63) but with coefficients constrained to be the same and negative. This suggests that when a specification such as (63) is estimated, it is possible to obtain negative coefficients for local variables even in the presence of dynamic agglomeration economies, and negative signs have indeed been obtained in the literature.

Taking all the intuitions in (61), (63) and (65) together, one may consider a specification with static and dynamic agglomeration effects (as we did for productivity in Section 2.2), as well as

time autocorrelation of residuals, which leads to:

$$\begin{aligned}\ln L_{c,s,t} - \ln L_{c,s,t-1} &= -\rho \ln L_{c,s,t-1} + \beta \left(\ln den_{c,t} - \ln den_{c,t-1}\right) \\ &+ \mu \left(\ln area_{c,t} - \ln area_{c,t-1}\right) + \vartheta \left(\ln spe_{c,s,t} - \ln spe_{c,s,t-1}\right) \\ &+ \widetilde{\beta} \ln den_{c,t-1} + \widetilde{\mu} \ln area_{c,t-1} + \widetilde{\vartheta} \ln spe_{c,s,t-1} + \varepsilon_{c,s,t},\end{aligned} \tag{66}$$

This specification involves time variations of static effects, dynamic effects, and inertia in industrial employment due to the time autocorrelation of residuals.[20] Rearranging terms to eliminate current and past specialisation (as their coefficients are not identified), we finally get:

$$\begin{aligned}\ln L_{c,s,t} - \ln L_{c,s,t-1} &= \tfrac{\widetilde{\vartheta}-\rho}{1-\vartheta} \ln L_{c,s,t-1} + \tfrac{\beta-\vartheta}{1-\vartheta} \ln den_{c,t} + \tfrac{\mu-\vartheta}{1-\vartheta} \ln area_{c,t} \\ &+ \tfrac{\widetilde{\beta}-\beta+\vartheta-\widetilde{\vartheta}}{1-\vartheta} \ln den_{c,t-1} + \tfrac{\widetilde{\mu}-\mu+\vartheta-\widetilde{\vartheta}}{1-\vartheta} \ln area_{c,t-1} + \varepsilon_{c,s,t}\,.\end{aligned} \tag{67}$$

which is a specification close to the one estimated by Henderson (1997) and Combes et al. (2004). Alternatively, one can replace past industrial employment $L_{c,s,t-1}$ by $den_{c,t-1}$ $area_{c,t-1}$ $spe_{c,s,t-1}$ to rather consider a specification with past specialisation although the same parameters are identified.

Unfortunately, the five coefficients in equation (67) are combinations of the seven parameters of interest. It is thus difficult to interpret the estimated coefficients even if one is able to deal with the endogeneity of right-hand side variables. For instance, a negative impact of past industrial employment is compatible not only with the presence of inertia in the series together with a positive static effect of specialisation, but also with a negative static effect of specialisation. Similarly, a positive impact of past local determinants is not incompatible with a negative impact of some static or dynamic agglomeration effects. As there are more parameters of interest than estimated coefficients, the different effects cannot be disentangled. The model could be augmented with other local characteristics such as market potential or diversity, and more lags of industrial employment, using statistical tests to determine how many lags should finally be kept. However, the same identification issues would remain as the impact of these variables would mix again static and dynamic effects.

Another point that we have not discussed so far about equation (60) is that local wage (or local wage growth if the dependent variable is employment growth) should be used as a control variable in the empirical specification if one wishes to restrict the interpretation of the effects of local characteristics to their role in $p_{c,s,t}$, $A_{c,s,t}$ and $r_{c,s,t}$ only (consistent with the analysis on productivity) and avoid considering their role in $w_{c,s,t}$. Since one estimates a labour demand equation, the local wage is expected to have a negative effect on local employment. For given wages, agglomeration effects increase labour demand, and therefore we expect a positive effect of

[20] This specification is not completely consistent with theory. It is possible to derive a specification which is consistent but it is much more intricate.

density, area and market potential among others on local employment as in the case of productivity.

However, controlling for wages means that only a partial equilibrium effect of agglomeration economies is captured. It corresponds to the direct impact of agglomeration economies on labour demand but it does not capture the feed-back effects on this demand resulting from the wage change induced by agglomeration. Moreover from the econometric point of view, controlling for wages raises serious additional endogeneity issues, on top of those described above when the dependent variable measures productivity.

One can choose not to control for local wage but then the impact of local characteristics on local employment operates not only through $p_{c,s,t}$, $A_{c,s,t}$ and $r_{c,s,t}$ but also $w_{c,s,t}$, and the effect through wage is negative. Typically, agglomeration economies raise nominal wages, which in turn yield a decrease in labour demand. The overall impact of agglomeration economies on employment is now ambiguous, and in particular it can be negative. On the one hand, agglomeration economies that increase $p_{c,s,t}$, $A_{c,s,t}$ and decrease $r_{c,s,t}$ tend to positively affect employment, on the other hand, they also increase $w_{c,s,t}$, which tends to negatively affect employment. When the effect of density on local employment is found to be negative, one does not know if density has a negative effect on productivity, and therefore a negative effect on employment because productivity is positively related to employment, or if density has a positive effect on productivity which in turn has a positive effect on wages themselves affecting employment negatively. For instance, Cingano and Schivardi (2004) get opposite signs for some of the common determinants of productivity and employment, based on the same Italian dataset. This suggests that a positive effect of agglomeration economies on local productivity can actually turn into a negative effect on local employment.

Finally, Combes et al. (2004) also propose to break down local employment into two terms, employment per firm and the local number of firms:

$$\ln L_{c,s,t} = \ln \left(\frac{L_{c,s,t}}{n_{c,s,t}} n_{c,s,t} \right) = \ln \frac{L_{c,s,t}}{n_{c,s,t}} + \ln n_{c,s,t}, \tag{68}$$

where $n_{c,s,t}$ is the local number of firms within the industry. One can evaluate separately the impact of local characteristics on average employment in existing firms and on the number of firms. Indeed, urbanisation and localisation variables can have different effects on the intensive and extensive margins of employment. In first differences, the analysis indicates whether agglomeration economies have the same or opposite effects on internal firm growth and on external growth, or whether effects are stronger for one or the other employment growth components. Finally, note that some authors evaluate the effect of local human capital on employment growth in the spirit of what has been done for productivity, as for instance Simon (2004) for the US, and Suedekum (2008) and Suedekum (2010) for Germany. The interpretation is again blurred by the existence of substitution effects between high-skill and low-skill workers as discussed in Subsection 3.3.

6.1.2 Total employment, specialisation, diversity, and human capital

The explanatory variables introduced into employment growth regressions are usually very similar to those considered in productivity regressions, except that local density is replaced by local total employment. Estimated specifications generally involve dynamic agglomeration effects following (63) but not static effects. Results on the effect of total employment on industrial employment growth clearly illustrate the diversity of results obtained in the literature on local employment growth. Beyond the fact that samples for different countries and periods are used, the previous section illustrates how the use of different specifications does change the interpretation of estimated effects. For instance, Combes (2000) finds for France that the local market size has a positive effect on industrial employment growth for manufacturing industries but a negative effect for services industries. Viladecans-Marsal (2004) finds for Spain that the effect on industrial employment is not significant for three out of six industries, while it has a bell-shaped effect in the three other industries. Blien et al. (2006), who extend Blien and Suedekum (2005), obtain for Germany that local market size plays a positive role on industrial employment growth both for manufacturing and service activities. There are two recent studies on Italy, one that pools together manufacturing and service industries (Mameli et al., 2008) and one that focuses on business services (Micucci and Giacinto, 2009). Both conclude that total employment has a positive impact on industrial employment growth.

As we mentioned above, the question of the spatial decay of agglomeration effects is crucial. For the US, Desmet and Fafchamps (2005) consider the impact on local employment growth of total employment and industrial employment share at various distances from the location. They show that for non-service industries, such as manufacturing and construction, the effects are negative for distances below 20 kilometres, but are slightly positive for distances between 20 and 70 kilometres. This is consistent with employment moving away from city centres with high aggregate employment to nearby locations. Service industries exhibit a different pattern for the effect of total employment: the coefficients are positive at distances below 5 kilometres, and slightly negative at distances between 5 and 20 kilometres. This is consistent with employment growing faster in city centres and more slowly in nearby areas. Unfortunately, this question has rarely been addressed for European economies. Viladecans-Marsal (2004) studies the role on industrial employment of the local characteristics of neighbouring cities in Spain. She finds the effects of total local employment and employment in neighbouring locations to be significant in two out of the six industries she considers. In the same vein, and still on Spanish data, Solé-Ollé and Viladecans-Marsal (2004) show that growth of the central municipality within metropolitan areas has a positive effect on growth in the suburbs. Micucci and Giacinto (2009) also find for Italy a significant impact of distant locations on local employment growth.

The impact of diversity on productivity has been found to be not robust and this is also true for its effect on industrial employment growth. Whereas Glaeser et al. (1992) find a positive impact

of diversity (measured by the share of the five largest industries within the city) on industrial employment growth, Henderson et al. (1995) who use a Herfindhal index over all local industries obtain a significant positive effect in a couple of high-tech industries only. For France, Combes (2000) finds that the same diversity index has also a positive impact on employment growth in service industries but a negative one in most manufacturing industries, although it is positive for a few of them. For Spain, Viladecans-Marsal (2004) finds a positive static effect on employment for three industries but a negative effect for some others and a non-significant effect for two of them. For Germany, Blien et al. (2006) find that diversity has a positive effect on employment growth in both manufacturing and service industries, the effect being strong in the manufacturing industry. Diversity is also found to have a significant positive impact in Italy according to Mameli et al. (2008).

The impact of specialisation is difficult to assess because its role on agglomeration economies cannot be disentangled from the mean reversion process of industrial employment as shown earlier. The impact of specialisation is found to be negative in both manufacturing and service industries in France by Combes (2000), in Germany by Blien et al. (2006), and in Italy by Mameli et al. (2008). This result may arise from strong mean reversion that more than compensates positive agglomeration effects. Van Soest et al. (2006) obtains a positive effect of specialisation in Netherlands but the impact is very local and dies out quickly with distance.

Glaeser et al. (1992) popularised the use of the local average size of firms in industry as a determinant of localisation economies as discussed in Subsection 3.2. Both Combes (2000) for France and Blien et al. (2006) for Germany find that the presence of larger firms reduces employment growth in both manufacturing and service industries. To refine the role of local firm size, Combes (2000) introduces a local Herfindahl index of firm size heterogeneity. He finds that the local concentration of employment within large firms is also detrimental to local growth. Therefore, in France, the local market structure that fosters employment growth the most appears to be small firms of even size. A further example of the difficulty of interpreting the findings of this literature is given by Mameli et al. (2008) who show from Italian data that the effect of most local determinants on local employment is not very robust, in the sense that their sign changes depending on the industrial classification which is used.

Finally, local human capital is found to positively affect total employment growth, both in the US by Simon (2004) and in Germany by Suedekum (2008). However, the latter study emphasises that mostly unskilled employment growth is favoured, which is consistent with the presence of strong substitution effects between the two groups of workers and weak agglomeration effects.

6.1.3 Dynamic specifications

A crucial question is the time needed for a determinant of agglomeration economies to have a sizeable effect. The availability of panel datasets has generated a series of papers that estimate

jointly the dynamics of both the dependent local variable and local determinants of agglomeration economies in specifications with multiple lags involving both static and dynamic agglomeration effects. In other words, instead of estimating specifications described in Subsection 6.1, researchers estimate full auto-regressive models, as initially proposed by Henderson (1997) for US cities. Once this kind of model is estimated, short-run effects of local determinants can be distinguished from their long-run effects.

For instance, Blien et al. (2006) show that in Germany the impact of diversity dies out quickly over time, in both the manufacturing and services sectors. This means that diversity has no long-run effects. Similarly, the effect of local firm size is significant in the short run but not in the long run in the two sectors. As mentioned above, Combes et al. (2004) propose decomposing industrial employment into average employment per firm and the number of firms in the local industry. They then estimate from French data a Vector Auto-Regressive model involving these two dependent variables (this approach has been replicated with German data by Fuchs, 2011). It is found that the local determinants of the growth of existing firms are not necessarily the same as those that promote the creation of new firms. Overall, there is a greater inertia in the adjustment process in the United States than in France and Germany. Lagged values stop being significant after one year of lag for France and Germany. This is starkly at odds with the six- or seven-year significant lags found in Henderson (1997) for the US.

Unfortunately, as emphasised in Subsection 6.1.1, interpretations of estimated coefficients in terms of static and dynamic agglomeration effects remain very difficult because both types of effect can enter each estimated coefficient. Moreover, even if the structure of Vector Auto-Regressive models makes them rather suited to deal with endogeneity concerns by using dynamic panel estimation techniques, the application of such techniques is debatable in the context of agglomeration effects as argued in Subsection 4.3.3. Ultimately the literature using dynamic specifications remains descriptive and is not really able to deliver causal interpretations of the effects in terms of agglomeration economies.

6.2 Firms' location choices

Rather than assessing the impact of local determinants of agglomeration economies on productivity or industrial employment, some authors have tried to evaluate the impact of these determinants on the location choices of firms. Firms should locate where their expected profit is the highest. As profit increases with productivity, the local determinants of productivity should also affect firm location choices. This is the intuition motivating the approaches presented in this subsection. They lead to applications usually relating to location choices of foreign direct investments (FDI) or determinants of firm creation.

6.2.1 Strategies and methodological concerns

To assess the role of local determinants of firm location choices, Carlton (1983) proposes to use the discrete choice modelling strategy developed by McFadden (1974). The idea is that, for any given firm, the value of each location depends on a deterministic local profit and an idiosyncratic component. The local profit is supposed to be the same for all firms but the idiosyncratic component varies across firms (and components are identically and independently distributed across locations for a given firm). This prevents firms from all choosing the same location, which would not correspond to reality. Assuming that idiosyncratic components follow extreme value laws, the firm location choice follows a logistic model, or Logit, which is quite easy to estimate.

Economic geography models predict how firms distribute themselves across space according to local profits, which are non-zero in the short run under imperfect competition. The location choice thus depends on the same quantities as those that enter the productivity equation (50) (the prices of goods and intermediate inputs, the technological level of the firm, and workers' efficiency) as well as the nominal wage. As a result, any of the urbanisation and localisation variables which enter the empirical specification of productivity can be included in a specification explaining firm location choices. However, interpretations are even more difficult than in the case of industrial employment, as there are direct and indirect effects which sometimes go in opposite directions. Indeed, profits depend not only on productivity but also on input use and output quantity which are themselves influenced by agglomeration effects but are not introduced in the regression. One can also choose whether or not to control for the local level of wages but interpretations then differ as in the case of industrial employment. Therefore, proposing correct and precise interpretations is difficult because many effects are at play and they interfere in non-linear ways to shape local profits.

Furthermore, almost all the local variables explaining location choices can be considered to be endogenous, precisely due to the location choices of both firms and workers which induce reverse causality affecting most local determinants of agglomeration economies. Unfortunately, this kind of issue is tackled even less often in empirical studies on firm location choices than in the literature on the local determinants of productivity and employment. At best, authors lag explanatory variables by one period of time, which is certainly not enough to correct for any endogeneity bias that may occur. To cope with the problem of omitted local variables, some authors include regional dummies at a geographical scale larger than the one considered for location choices, while others exploit time series and introduce local fixed effects. The same important caveats appear as for productivity studies and they are detailed in Subsection 4.3.

For all these reasons, the literature on firm location choices has to be considered as mostly descriptive. A safer route to assess the role of agglomeration effects on firm location choices would probably be to consider much more structural approaches, which however present the drawback of considering a more limited number of agglomeration channels.

Besides these limits, it is possible to enrich the approach when studying the location choices of firms among places in several countries using a nested logit model involving several stages. For instance, firms first choose the country where to locate and then, conditional on this choice, choose the region or city within the country. Two additive random components are now considered, one specific to the region and one specific to the country, and they are assumed to be independent. This structure produces a total random component correlated between regions within a given country, and the correlation can be estimated simultaneously with the other parameters in the model. In fact, the effects of local determinants of location choices at the different spatial scales are evaluated separately, once the geographical decomposition of the whole territory has been chosen (for instance, countries or continents, divided themselves into regions or cities). The nested logit approach has the advantage of limiting the number of possible locations considered for a firm choice at a given stage. This can be a desirable feature considering current computer capacities, especially if some fixed effects (for industries or other geographical scales) are introduced in the model. These estimation strategies have been considered in empirical studies that take either a reduced form approach, such as Carlton (1983), or a more structural approach where firm location choices are part of an economic geography model, such as Head and Mayer (2004).

Research based on discrete location choice models has primarily been applied to FDI because the determinants underlying their location decisions are more discernible than those of domestic firms that are less footloose. In particular, location choices are made by multinational firms in a relatively short period of time, without bearing the weight of historical contingencies like national firms. This makes them more appropriate candidates to test for the presence of agglomeration effects. An alternative approach adopted in a number of papers consists of considering the number of firm entries in a region as the dependent variable, and studying its determinants with a simple Tobit approach, or a count model such as the Poisson or the negative binomial, or even with a linear model. The Tobit model takes into account the left censorship of the dependent variable but considers that this variable is continuous. The main advantage of count models is that there is no computational limit on the number of alternatives such as in the Logit model. However, there are strong distributional assumptions on residuals. The standard linear model does not impose any assumption on the distribution of residuals and is very flexible for the number of covariates that can be considered but it ignores the discrete nature of the data and left censoring.

6.2.2 Discrete location choice models

Among early studies on the effect of local economy characteristics on location choices of FDI, Head et al. (1999) focus on the determinants of firm location choices between the 50 states of continental US, while Guimaraes et al. (2000) conduct a similar exercise for the 275 regions in Portugal, which are much smaller. Because of the urban and regional perspective of our survey, we do not discuss studies on location choices between countries. It may be noted, however, that their findings do not

significantly differ from those on location choices within a country even if the nature of underlying agglomeration economies is likely to differ.

As predicted by theory, the first factor that is almost systematically found to play a positive role on location choices of FDI is the size of the local economy. For instance, market size is measured with local total income in Head et al. (1999), and with two variables, manufacturing and services employment, in Guimaraes et al. (2000). Among other determinants of firm location choices is market access. Guimaraes et al. (2000) consider the distance to the main cities of Portugal as a proxy. At the European level, Head and Mayer (2004) compare the performance of Harris and structural market potential variables in explaining the location choices of Japanese affiliates across European regions at the NUTS 2 level. They find that both have a significant positive impact on these choices, even when controlling for a substantial number of other variables. Basile et al. (2008) analyse the location choices of multinational firms of various nationalities in 50 regions in eight EU countries. External market potential is found to have a significant positive effect as well as the own region total value added which is considered simultaneously. However, both effects appear to be mainly driven by location choices of European multinationals, and they are not significant for non-European ones.

The positive impact of market potential seems to be fairly universal and it is confirmed when data are disaggregated along various dimensions. For instance, Crozet et al. (2004) find a positive effect on FDI in France whatever the country of origin of firms. When studying FDI in Germany, Spies (2010) always find a positive effect of market potential when conducting estimations for each industry separately. Pusterla and Resmini (2007), who focus on FDI in the NUTS 2 regions in four Eastern European countries, find that both local manufacturing employment and market potential variables positively affect FDI, although most of the impact is on low-tech industries and not on high-tech ones.

As in the literature on productivity determinants, the functional form chosen for the role of distance in the market potential -the inverse of distance in most cases- assumes a fast spatial decay of agglomeration effects. The role of proximity has been further investigated. Basile (2004) for instance finds a negative effect on FDI of agglomeration in adjacent provinces in Italy, while at the same time agglomeration in own province has a positive effect. Interestingly, the authors are able to distinguish between foreign acquisitions and greenfield investments. The effect of the local number of establishments is found to be significantly positive only for foreign acquisitions. However, local demand measured by electricity consumption, which is also introduced into the specification, has a positive influence on the two types of firms. Greenfield investments are more appealing for evaluating the role of agglomeration effects because firms have more freedom in their location choices.

This literature almost systematically considers the role of a variable absent from local productivity or growth estimations: past foreign presence in the region. This variable can have effects

going in opposite ways. On the one hand, it may attract future FDI because it reflects unobservable characteristics of the region that are also beneficial to new FDI, or because it reflects an existing business network that may be useful to new FDI. On the other hand, past foreign presence may have a negative impact on new FDI because of competition effects. From a theoretical point of view, it is also difficult to assess how such a variable interferes with other local determinants of agglomeration economies, in particular the size of the local economy. As always, absent relevant instruments and natural experiments, identifying causal effects is very difficult.

Current FDI is shown to be positively correlated with previous FDI. For instance, past FDI is found to attract Japanese affiliates in European regions (Head and Mayer, 2004), and to induce both acquisitions and greenfield investments in Italy (Basile, 2004). Past investment also has an influence in both low- and high-tech industries in Germany (Spies, 2010), Eastern European countries (Pusterla and Resmini, 2007), and Ireland (Barrios et al., 2006). Basile et al. (2008) find for European regions a positive effect of foreign presence on both European and non-European FDI. Crozet et al. (2004) study FDI in France by country of origin and find a positive effect of past presence for specific countries only, the largest effects being observed for Japan, the UK, Belgium, and the US. Finally, Devereux et al. (2007) find a positive effect of past foreign investment in the UK on both new investment by domestic firms and FDI, the effect being larger for FDI. The role of social and business network has also been indirectly investigated through variables such as the distance to home country or headquarters, which is found to have a negative impact on FDI in France by Crozet et al. (2004) and on European FDI in European regions by Basile et al. (2008). Generally, sharing a common language also has the expected positive effect on FDI, and this can be interpreted as indirect evidence of the presence of communication externalities.

As for productivity, authors also study the effect of local industry characteristics on location choices. FDI is fairly systematically found to be positively correlated with specialisation, usually measured by the local count of domestic firms in the industry at the European level (Head and Mayer, 2004), or within countries such as in Portugal (Guimaraes et al., 2000), France (Crozet et al., 2004) or the UK (Devereux et al., 2007). Devereux et al. (2007) also find a positive impact of local industrial diversity. For Ireland, Barrios et al. (2006) find that diversity has had a significantly positive impact on FDI since the 1980s, but not before, and only for high-tech firms for which specialisation has no impact. Conversely, whereas diversity does not matter for low-tech firms, specialisation has a positive impact on low-tech FDI. Hilber and Voicu (2010) find for Romania that both domestic and foreign industry-specific agglomeration measures positively affect FDI, but only the effect of domestic agglomeration is robust to the introduction of regional fixed effects. The same is found for the effect of domestic industry-specific agglomeration in neighbouring regions. The positive effect of diversity that is estimated without regional fixed effects is found to be not robust to their introduction.

Guimaraes et al. (2000) distinguish between the impact of manufacturing and service concen-

tration, and find a larger impact from service concentration. This result has been confirmed in later studies, in particular for Eastern European regions. According to Cieślik (2005), service concentration has a significant positive large effect on FDI in Poland at the NUTS 3 level (49 regions), and the same is found for Romania at the NUTS 3 level (21 regions) by Hilber and Voicu (2010), even when region fixed effects are included in the specification. As an example, an increase of 10.0% in the density of service employment in a Romanian region makes the average Romanian region 11.9% more likely to attract a foreign investor.

As we can see, there are a variety of results that emphasise effects going more or less in the same direction but that remain difficult to compare (because authors usually estimate different specifications) and interpret (because of both the large number of possible effects and the possible presence of reverse causality).

These issues are even more important when studying the role of local labour markets in FDI as has been done in the literature. In particular, the impact of local labour costs has been investigated but a significant concern is that authors are rarely able to control simultaneously for the local quality of labour. The labour cost per efficient unit of labour would be predicted by theory to influence location choices but only the nominal cost is in general available. When labour efficiency is not taken into account, a positive impact of wages on the choice of a location may reflect the presence of high-skill workers. Moreover, wages are simultaneously determined with firm location choices, and this endogeneity issue is usually not addressed. The endogeneity issue may be even more important when the local unemployment rates are introduced into the specification and micro-foundations of the specification are even more unclear. A high local unemployment rate may reflect a large labour supply, and thus low wages or, on the contrary, wages that are too high and cause unemployment. Ultimately, due to the lack of theoretical background for empirical specifications, we think that little can be learnt from the impact of these variables. This is why we do not detail here their estimated effects, and we believe that a better use of theory will be required to really investigate the role of local labour markets.

6.2.3 Firm creation and entrepreneurship

Some recent literature argues that the location choices of new entrepreneurs and their determinants are worth studying because they should be more informative on the role and magnitude of agglomeration effects than the location choices of new plants by existing firms, as these choices are influenced by the locations of existing establishments of these firms. Unfortunately, as pointed out by Glaeser et al. (2010b), the literature on this topic is relatively small. Some contributions relate to the literature on innovations, and are therefore surveyed in Carlino and Kerr (2015). We describe here some contributions that describe the determinants of firm creations in a more general way.

Among papers on the US, Rosenthal and Strange (2003) show that firm creation is more

important when the own-industry employment located within the first mile is larger, but the effect then vanishes rapidly with distance. Indeed, the impact within the first mile is 10 to 1000 times larger than the impact 2 to 5 miles away. They do not find any robust impact of urbanisation on firm creation. Glaeser and Kerr (2009) propose disentangling among plant creations those that do not result from existing firms, as it is a better measure of entrepreneurial activity. The local level of activity appears to favour entrepreneurship, as it goes along with the presence of many small local suppliers. Glaeser et al. (2010a) find not that there are higher returns where entrepreneurs settle but that entrepreneurs rather choose places where there are larger local entrepreneurial pools. Using the same dataset, and in the spirit of papers on determinants of local industrial employment, Delgado et al. (2010) augment the specification with dynamic effects and argue that mean reversion effects co-exist with agglomeration gains.

Among contributions on other countries, Figueiredo et al. (2002) investigate the location choices of entrepreneurs in Portugal. Interestingly, they are able to distinguish between native and non-native entrepreneurs, and agglomeration effects are found only for non-natives. At a fine geographical scale, Arauzo-Carod and Viladecans-Marsal (2009) show for Spain that firm creation increases with own-industry previous entries. The effect is larger, the higher the technological level of the industry. Finally, Harada (2005) and Sato et al. (2012) find for Japan that a larger market size increases the willingness to become an entrepreneur, and that the effect is U-shaped for the share of them that become entrepreneurs eventually. Put differently, people are more often entrepreneurs in both large and small locations. By contrast, Addario and Vuri (2010) find that population density reduces the probability of being an entrepreneur in Italy even if entrepreneurs' earning are larger in denser areas.[21]

Overall, there is a great variety of results, which may be related to the estimation of different specifications and the way endogeneity issues are handled, especially as these issues are not always addressed. Still, once the burgeoning literature on location choices of entrepreneurs is better related to theory, and takes better into account spatial sorting and reverse causality, it should deliver interesting conclusions on the local determinants of entrepreneurship.

7 Identification of agglomeration mechanisms

The literature assessing the effects of local determinants of agglomeration economies on local outcomes estimates the overall net impacts of local variables, but it does not enter the black box of the underlying mechanisms at stake. Some attempts to identify some of these mechanisms have been made recently in three directions. A series of papers focuses on job search and matching effects, and evaluates whether agglomeration effects on productivity are related to the way local labour markets operate. Other authors have taken an indirect route by testing whether industrial

[21] There is also a recent literature on developing countries (see Ghani et al., 2013, 2014).

spatial concentration or firms co-location relates to industry characteristics associated to the Marshallian three broad families of agglomeration mechanisms: labour pooling, knowledge spillovers, and input-output linkages. Last, a couple of case studies have been proposed to quantify specific agglomeration effects.

7.1 Labour mobility, specialisation, matching, and training

Some of the gains from agglomeration arise from an increase in job mobility and better matching between workers and firms. Some studies assess whether agglomeration increases the frequency of workers' moves between firms, industries, or occupations, as well as the chances for the unemployed of finding a job. Freedman (2008) studies the effect of specialisation on workers' job mobility and earnings dynamics for the software publishing industry in one anonymous state using a US longitudinal matched employer-employee dataset. Higher specialisation in a 25km radius increases the chances of moving between two software jobs. A wage regression also shows that specialisation within a 25km radius lowers the initial wage but is also associated with a steeper wage profile leading to a wage premium.

Using the National Longitudinal Survey of Youth, Wheeler (2008) evaluates the effect of local population, density, and diversity on mobility between industries depending on the number of previous job moves. When looking at a sample of first job changes, he finds that industry changes occur more often in large and diverse local markets than in small and non-diversified ones. Once several jobs have been held, the positive relationship becomes negative. As workers in large markets also tend to experience fewer job changes overall, the evidence is consistent with agglomeration facilitating labour market matching. In a similar spirit, Bleakley and Lin (2012) study the effect of the metropolitan area employment density on occupation and industry changes using US data. They instrument current local density with historical local density and current density at the state level. The rate of transitions of occupation and industry is found to be lower in denser markets but the result is reversed for younger workers, which is consistent with Wheeler (2008)'s interpretation. The local employment share in own industry or own occupation also has a negative effect on industry and occupation changes.

The effects of agglomeration variables on the job search process is investigated by Di Addario (2011) for Italy. She estimates the effects of local population and specialisation on the probabilities for non-employed individuals of searching for a job and getting employed. Agglomeration variables are instrumented with historical population, seismic hazard, and soil characteristics. Overall, results show that a larger local population and location in an industrial district or super-district increase the probability of being employed. Conversely, the impact of any variable on search behaviour is found to be zero.

Some authors have investigated whether matches between workers and firms are more productive in larger/denser areas. Some approaches used to evaluate the effect of matching on produc-

tivity in a static framework are discussed in Subsection 2.3. In an application, Wheeler (2006) finds that wage growth is more important in large cities than small ones and that this difference is mostly related to differences in wage growth when changing jobs. This is consistent with better matching in larger cities. However, this study does not take into account the endogeneity of job and location mobility. This can be done using a more structural approach as explained in Subsection 2.4. Baum-Snow and Pavan (2012) estimate a structural model and find that match quality contributes little to the observed city size premium, in comparison with other static and dynamic agglomeration effects. Differences in conclusions may be due to differences in the structure of static and dynamic models, and more specifically how the endogeneity of individual choices is handled.

Alternative static approaches have been proposed to assess the role of match quality. Andersson et al. (2007) use matched worker-firm panel data on California and Florida to estimate a wage equation involving worker and firm fixed effects. They then compute for each county the correlation across firms between the firm fixed effect and the average worker fixed effect within the firm. The correlation is regressed at the county level on average firm fixed effect, average worker fixed effect, as well as density. The estimated coefficient of density is found to be positive and significant, indicating improved matching in denser areas. Figueiredo et al. (2014) evaluate the effect of density on matches between workers and firms using Portuguese employer-employee panel data. Their empirical strategy has two stages. First, they estimate a wage equation involving worker, firm and match effects. Second, estimated match effects are regressed on explanatory variables including in particular density and specialisation, as well as worker and firm fixed effects. The estimated effect of density in second stage is not significant. The effect of specialisation is significantly positive at the 10% level only. What remains unclear is to what extent the sole match effect captures all complementarity effects between workers and firms. Wage is expressed in logarithm in the first-stage specification, which means that the exponentiated product of worker and firm fixed effects also captures complementarities.

Finally, Andini et al. (2013) assess for Italy whether there is an effect of density (and classification into an industrial district) on worker and firm individual measures of labour pooling. Density is measured at the local labour market level and is instrumented using historical values. Individual outcomes are the change of employer or type of work, or both, workplace learning, past experience, training by the firm, skill transferability, difficulty of replacing the worker or finding another job, measures of specialisation and the appropriateness of experience and education. Firm outcomes are the share of terminations that are voluntary, the share of vacancies filled from workers previously employed in the same industry, the number of days to train key workers, a measure of appropriateness of a new worker in terms of education and experience. Overall, results support theories of labour pooling, but evidence is weak, possibly due to the small size of the datasets. In particular, there is some evidence of a positive effect of agglomeration on turnover, on-the-job training, and improvement of job matches.

Another possible mechanism that might lead to higher productivity in cities is task specialisation. The underlying idea is that there are benefits to the division of labour, and this division is limited by the extent of the market. The division of labour is then expected to be greater in larger markets. There are a few bits of research on the relationship between the division of labour and city size. Duranton and Jayet (2011) study this relationship using information on more than 5 million workers in 454 occupations and 114 sectors extracted from the 1990 French census. It is shown that even after taking into account the uneven distribution of industries across cities, larger cities exhibit a larger share of workers in scarcer occupations. For example, the difference between Paris and the smallest French cities is around 70%. For Germany, Kok (2015) shows that the specialisation of jobs and the required level of cognitive skills increase with city size. To our knowledge, the links between city size, the division of labour and productivity have not yet been investigated.

Last, some authors have investigated whether knowledge spillovers arise from the mobility of workers between firms within the same local labour market. Serafinelli (2014) shows that in the region of Veneto, Italy, hiring a worker with experience at highly productive firms significantly increases the productivity of other firms. According to his results, worker flows would explain around 15% of the productivity gains experienced by other firms when a new highly productive firm is added to a local labor market. Combes and Duranton (2006) propose a model in which firms choosing their location anticipate that they can improve their productivity by poaching workers from other firms. However their workers can be poached too unless they are paid higher wages, which makes their production costs larger. Some authors have proposed to test this story indirectly by studying how training within firms varies across city size, the alternative to training being to poach workers who are already trained from other firms. Brunello and Gambarotto (2007) for Italy, Brunello and Paola (2008) for the UK, and Muehlemann and Wolter (2011) for Switzerland, show that indeed there is less on-the-job training in larger markets, and this is particularly true in the UK.

Overall, the literature on mobility, job search, and training comprises interesting attempts to determine the agglomeration mechanisms that relate to the labour market. It remains mostly descriptive though and would gain from considering approaches more grounded in theory and tests of particular models.

7.2 Industrial spatial concentration and co-agglomeration

Another strand of the literature has tried to identify the separate role of the three main types of mechanisms underlying agglomeration economies according to Marshall (1890): knowledge spillovers, labour pooling, and input/output linkages. For that purpose, a couple of papers augment the specifications of employment or firm creation presented in Section 6 with variables that should capture these three types of mechanisms. A larger number of papers, which we present

first, compute spatial indices of concentration or co-agglomeration for every industry, and then regress them on industry characteristics related to the three families of mechanisms. As analyses usually do not rely on a precise theoretical framework, this literature is for the moment mostly descriptive.

Kim (1995) is among the first to compute a spatial concentration index for some industries, in his case the Gini spatial concentration index (see Combes et al., 2008b), and regress it on industry characteristics and more particularly on average firm size. His purpose is to test the intuition that industries with stronger increasing returns to scale, which should be characterised by larger firms in equilibrium, are spatially more concentrated. The spatial concentration index is computed for a division of the US into 9 large regions, for 20 industries, and for 5 points in time over the 1880-1987 period. The share of raw materials in production is introduced in the specification supposedly to control for the impact of comparative advantages on spatial concentration, and industry fixed effects are used to capture the role of industry effects constant over time.

There are major limitations to this kind of empirical strategy. Even simple economic geography models show that increasing returns to scale interact with trade costs and the degree of product differentiation to fix the degree of spatial concentration in equilibrium (see Combes et al., 2008b). However, only one industry characteristic among these three is introduced in the specification. It is thus necessary to make the strong assumption that either the two other characteristics are not correlated with the first one or they are sufficiently invariant over time to be captured by industry fixed effects. If trade costs and product differentiation indices were available, considering them in the specification would certainly not be straightforward since theoretical models usually predict highly non-linear relationships between outcomes and underlying parameters. Introducing these characteristics as additional separate linear explanatory variables could be too extreme a simplification. Similarly, comparative advantage theory stresses the role of the interaction between factor intensity in the production function and regional factor endowments. Controlling for factor intensity but not for the distribution of endowments over space leads to ignoring the mechanism that generates regional specialisation. Lastly, some mechanisms affecting spatial concentration, such as knowledge spillovers and labour pooling, are not taken into account either.

Further studies have tried to assess the role of additional agglomeration mechanisms by augmenting the estimated specification.[22] Rosenthal and Strange (2001) is an interesting attempt in this direction. The spatial concentration measure is the Ellison and Glaeser (1997) index computed for 4-digit manufacturing industries in the US. Variables for the three types of mechanisms are considered. Input sharing is measured by the shares of manufacturing and non-manufacturing inputs in shipments. Knowledge spillovers are captured by innovations per dollar of shipment. Alternatively, some other authors also use R&D expenses. Measures of labour pooling are the value of shipments less the value of purchased inputs divided by the number of workers, the share

[22] They also use more detailed data, albeit on a shorter period of time.

of management workers, and the share of workers with at least a bachelor degree. These measures remain far from the intuition that industries with specific needs for some labour skills gain more than others from concentrating. A number of other control variables are introduced, many of which relate to primary input use with the purpose of capturing again comparative advantage effects. As only cross-section data are available, industry fixed effects can be introduced only at the 3-digit level and not the 4-digit level. The Ellison and Glaeser index takes into account in its construction an index of productive concentration that closely relates to the industry average plant size. Therefore it is not clear whether or not one should control for firm size, and the authors choose to leave it out of the specification.

Results obtained by Rosenthal and Strange (2001) are typical of this kind of study. Whereas labour pooling has a positive effect, knowledge spillovers have a positive impact on spatial concentration only when they are measured at a small scale (the zipcode). Reliance on manufactured inputs affects agglomeration at the state level but not at a smaller scale. By contrast, reliance on service inputs has a negative effect on agglomeration at the state level. Overman and Puga (2010) propose an alternative indirect measure of labour market pooling. It is based on the assumption that a labour pool of workers with adequate skills allows firms to absorb productivity shocks more efficiently. Using UK establishment-level panel data, they construct an establishment-level measure of idiosyncratic employment shocks and average it across time and establishments within the industry. They find that industries that experience more volatility are more spatially concentrated.

Long ago, Chinitz (1961) suggested that examining the degree of co-agglomeration of industries depending on their characteristics is another way to test for the presence of agglomeration economies. This approach is implemented in a systematic way by Ellison et al. (2010) who study the extent to which US manufacturing industries locate close to one another. The idea is to compute an index of co-agglomeration between two industries and to regress it on measures of proximity between the two industries in terms of labour pooling, knowledge spillovers, and input/output linkages. Labour pooling is measured with the correlation of occupation shares between the two industries. Alternatively, some authors use a measure of distance between the distributions of these shares in the two industries. Input and output linkages are proxied by the share of input from the other industry and the share of output to the other industry, respectively. Technological proximity is measured by two types of variables. The first type uses the shares of R&D flowing to and from the other industry. The second type uses patent citations of one industry made by the other industry. Such variables are in general not symmetrical. For instance, one industry can cite more the other than the opposite. Therefore, it is the maximum value of the variable for the two industries that is used in the regressions.

Importantly, in order to control for comparative advantage effects, the authors introduce among the explanatory variables a co-agglomeration index of spatial concentration due to natural advantages, which is an extension of the natural advantages spatial concentration index proposed by

Ellison and Glaeser (1999). Results are also provided for alternative co-agglomeration indexes. Indeed, a standard index such as the one of Ellison and Glaeser considers a classification of spatial units across which the economic activity is broken down and measures the concentration in these units. A limitation is that the relative location of units and the distances that separate them are not taken into account. As a result, the index is invariant up to any permutation of the units. For instance, it takes the same values if one relocates all units with large amounts of activity close to the centre of the economy or if one locates them at the periphery. Alternative measures of spatial concentration and co-agglomeration have been developed by Duranton and Overman (2005) to deal with this issue. They are based on the distribution of distances between establishments and can be computed for any spatial scope. One can assess whether there is concentration for a distance between establishments of 5 miles, 10 miles, and so on. Ellison et al. (2010) also estimate their specifications using the Duranton and Overman index computed for a distance of 250 miles. Finally, since explanatory variables are computed from the same quantities as the dependent variable, there might be endogeneity issues, and the authors propose to instrument explanatory variables with similar variables constructed from UK data instead of US data.

Results give some support to the three types of agglomeration mechanism. The largest effect is obtained for input-output linkages, followed by labour pooling. Kolko (2010) conducts a similar exercise for both manufacturing and service industries, using as additional measures of the links between industries variables related to the volume of inter-industry trade. He studies both agglomeration and co-agglomeration at various spatial scales: zip-code, county, metropolitan area, and state. Limitations are that he does not use distance-based concentration indices such as the Duranton and Overman index, he does not control for spatial concentration due to natural advantages, and he does not deal with endogeneity issues using instrumentation. Ultimately, trade between industries appears to be the main driver of industry co-agglomeration for both manufacturing and services. More precisely, service industries that trade with each other are more likely to co-locate in the same zip code, though not in the same county or state; by contrast, manufacturing industries that trade with each other are more likely to co-locate in the same county or state but not in the same zip code. Input sharing also positively affects co-agglomeration for both manufacturing and services and at any spatial level, and it is true for occupational similarity to some extent as a positive effect is found but only for services and at the zip-code level. As regards spatial concentration, labour pooling is the only variable having a significant impact. Its effect is positive but occurs in the manufacturing sector only.

Kerr and Kominers (2014) further study the determinants of spatial concentration in the spirit of Ellison et al. (2010). They compute the Duranton and Overman spatial concentration index for different industries and different distances. Values are pooled together and the resulting two-dimension panel is then regressed on dummies for distances interacted with an industry measure of knowledge spillovers, and then alternatively an industry measure of labour pooling. The proxies

used for these determinants are slightly different from those in other studies. As regards knowledge spillovers, the authors consider the citation premium for 0-10 miles relative to 30-150 miles. Labour pooling is captured by a Herfindahl index of occupational concentration computed over 700 categories. Most estimated coefficients obtained for interactions with dummies for distances decrease with distance, and they are significantly different from zero for short distances only. This suggests that establishments in industries with shorter knowledge spillovers or more labour pooling are more concentrated. Similar results are obtained whether one computes measures of knowledge spillovers and labour pooling using US data or UK data. Nevertheless, estimations for these two channels of agglomeration economies are conducted separately without confronting them in a single regression. Finally, estimated coefficients for interactions between dummies for distances and dependency to natural advantages tend to increase with distance and are significant for large enough distances only. This is consistent with the intuition that industries more dependent on natural advantages are more dispersed.

A difficulty faced by this literature is that the dependent variable is a complex function of certain quantities, such as local industrial employment, which relate to the quantities describing firms and establishments within the industry that are used in the construction of explanatory variables. Therefore it is not easy to argue about expected effects of explanatory variables in equilibrium, and this makes interpretations difficult. In light of this difficulty, Dumais et al. (1997) in a section not included in Dumais et al. (2002) propose re-examining the literature on industrial employment in order to assess the role of some specific agglomeration channels. They consider a specification where industrial employment is used as the dependent variable instead of an index of spatial concentration in the industry. Proxies for Marshallian externalities are constructed at the local level using the following strategy. Measures of proximity between industries as regards knowledge spillovers, labour pooling, input and output linkages are computed at the national level. For a given type of agglomeration channel, the local variable for an industry is then computed as the sum over all other industries of their proximity weighted by the share of these industries in the location. These local variables are also sometimes interacted with some of the local determinants of industrial employment presented in Subsection 6.1. All these terms serve as explanatory variables in the specification of local industrial employment.

Recently, a similar strategy has been implemented by Jofre-Montseny et al. (2011) to determine the effects of the different types of agglomeration economies on the location of new firms in Spain at the municipality level and city level.[23] In the same vein, Jofre-Montseny et al. (2014) estimate from Spanish data, for each industry separately, a firm location model with two main local explanatory variables, local employment within the industry and in other industries. The industry-specific estimates for these two variables are then regressed on industry characteristics

[23]Papers using the same strategy but for the study of agglomeration economies on TFP include Rigby and Essletzbichler (2002), Baldwin et al. (2010), Drucker and Feser (2012) and Ehrl (2013).

proxying for knowledge spillovers, labour pooling, input sharing, and energy and primary input use. We emphasised above the difficulty in interpreting estimates of employment growth specifications, while the authors propose here to further extend these specifications by introducing interactions between local determinants and factors influencing the different agglomeration forces at the industry level. Such extended empirical frameworks are necessarily even more ambiguous and difficult to interpret than the basic employment growth specifications that we discussed in Section 6.1.

Overall, this strand of literature is an interesting effort to identify the mechanisms underlying agglomeration economies. Ultimately though, it is very difficult to give a clear interpretation of the results, and conclusions are mostly descriptive. This is due to the weak links between estimated specifications and theoretical models. Another concern is whether the right measure of concentration or co-agglomeration has been chosen. The exact properties of concentration indices, even measures *à la* Duranton and Overman (2005), still need to be established. Moreover, one needs to assume that industry characteristics used as explanatory variables really capture the mechanisms they are meant to, and have additive linear effects whereas this is not granted. For instance, according to theory, two industries sharing inputs have more incentive to co-locate when trade costs for these inputs are large. In that perspective, variables capturing input/output linkages should be interacted with a measure of trade costs but this is not done in the literature. Finally, there are probably some endogeneity issues since the dependent variable and the explanatory variables are usually computed from the same quantities. However, the presence and channels of endogeneity are difficult to assess, and it is hard to conclude that some instruments are valid, as estimated specifications have usually not been derived from any precise theoretical framework. On the other hand, since the overall impact of agglomeration on productivity can be evaluated with reasonable confidence nowadays as we emphasised in previous sections, we think that investigating the relative magnitude of agglomeration channels is an important and promising avenue for future research. Descriptive evidence presented in this subsection could be used to build theoretical models from which specifications could be derived allowing the identification of agglomeration channels and strategies to tackle endogeneity concerns. Structural approaches applied to case studies, which are presented in the next subsection, constitute some first steps in that direction.

7.3 Case studies

Some specific mechanisms of agglomeration economies can be assessed through case studies of firms or industries for which the nature of possible density effects are known and can be specified.

An interesting structural attempt to evaluate the importance of agglomeration economies in distribution costs is proposed by Holmes (2011). The study focuses on the diffusion of Wal-Mart across the US territory and considers the location and timing of the opening of new stores. These new stores may sell general merchandise and, if they are super-centres, they may also sell food.

When operating a store, Wal-Mart gets merchandise sales revenues but incurs costs that include not only wages, rent and equipment costs, but also fixed costs. These fixed costs depend on local population density as well as the distance to the nearest distribution centre of general merchandise and, possibly, the distance to the nearest food distribution centre. Higher store density usually goes along with shorter distance from distribution centres. When opening a new store, Wal-Mart faces a trade-off between savings from a shorter distance to distribution centres and cannibalisation of existing stores. The estimation strategy to assess the effects of population density and proximity to distribution centres is the following. The choice of consumers across shops is modelled and demand parameters are estimated by fitting the predicted merchandise and food revenues with those observed in the data. An inter-temporal specification of Wal-Mart profit function taking into account the location of shops is then considered. In particular, this function depends on revenues net of costs which include wages, rent and equipment costs, as well as fixed costs. For a given location of shops, net revenues can be derived from the specification of demand where parameters have been replaced by their first-stage estimators. To estimate parameters related to fixed costs, the author then considers the actual Wal-Mart choices for store openings as well as deviations in which the opening dates of pairs of stores are reordered. Profit derived for an actual choice of store openings must be at least equal to that of deviations. This gives a set of inequalities that can be brought to the data in order to estimate bounds for the effects of population density and distance to distribution centres. It is estimated that when a Wal-Mart store is closer by one mile to a distribution centre, the company enjoys a yearly benefit that lies in a tight interval around 3,500 dollars. This constitutes a measure of the benefits of store density.

The benefits from economies of density in agriculture related to the use of neighbouring land parcels are evaluated by Holmes and Lee (2012). When using a particular piece of equipment, a farmer can save on set-up costs by using it across many fields located close to each other. Moreover, if a farmer has knowledge about a specific crop, it is worth planting that crop on adjacent fields, although this may be at the expense of reducing the crop diversity that can be useful against risks. The authors focus on planting decisions in the Red River Valley region of North Dakota for which there are a variety of crops and years of data on crop choice collected by satellite. More precisely, the focus is on quarter sections which are 160-acre square parcels. These sections can be divided into quarters of forty acres, each designed as a field. The empirical strategy relies on a structural model where farmers maximise their inter-temporal profit on the four quarters of theirs, choosing for each quarter the extent to which they cultivate a given crop (rather than alternative ones). Production depends on soil quality and the quantity of investment in a particular kind of equipment useful to cultivate the specific crop but which has a cost. It is possible to show that because of density economies arising from the use of the specific piece of equipment on all quarters, the optimal cultivation level for a crop on a quarter depends not only on the soil quality of this quarter but also on that of the other quarters. The specification can be estimated and parameters

can be used to assess the importance of economies of density. The authors find evidence of a strong link between quarters of a same parcel. If economies of density were removed, the long-run planting level of a particular crop would fall by around 40 percent. The authors also find that two-thirds of the actual level of crop specialisation can be attributed to natural advantages and one-third to economies of density.

8 Conclusion

Most of the literature identifies the overall impact of local determinants of agglomeration economies, but not the role of specific mechanisms that generate agglomeration effects. This is already a crucial element when assessing the role of cities. Major progress has been made in dealing with spatial sorting of workers and firms as well as endogeneity issues due to missing variables and reverse causality, especially when assessing the effect of density on productivity.

We developed a consistent framework that encompasses both the early attempts to estimate agglomeration effects using aggregate regional data and more sophisticated strategies using individual data, recently including some structural approaches. This allowed us to discuss most empirical issues and the solutions that have been proposed in the literature. We also presented the attempts to study the determinants of other local outcomes, namely employment and firm location choices, but more investigations are still needed. For instance, further theoretical and empirical clarifications would be useful when studying the determinants of local employment in order to better disentangle the short-term dynamics from long-term effects, and the respective role of labour demand and supply. The determinants of firm location choices have benefited so far from a very limited treatment of selection and endogeneity issues. Surprisingly, the impact of agglomeration economies on unemployment has received little attention and deserves more work at least from a European perspective as regional disparities in unemployment rates there remain large. Finally, identifying the channels of agglomeration economies is also clearly important but the related literature remains limited except for some contributions on innovation that are surveyed in Carlino and Kerr (2015). Meaningful strategies relying on sound theoretical ground to provide an empirical assessment of channels of agglomeration economies are still needed, and current evidence while being interesting is rather descriptive.

Some researchers have started to investigate routes complementary to those mentioned in our chapter. First, the existence of a spatial equilibrium implies that agglomeration costs are a necessary counterpart of agglomeration gains. This prediction is supported by Gibbons et al. (2011) who show that in Great Britain there is an almost one for one relationship between local housing costs and nominal earnings, which are higher in larger cities, once the effects of housing quality and workers skills are netted out. Therefore, assessing the presence of agglomeration costs can be used to provide indirect evidence for the existence of agglomeration gains. Second, some authors

have gone a step further by looking at the implications in terms of welfare of the simultaneous presence of agglomeration costs and gains. However, some effects have not yet been considered in the analyses, whereas they have some importance in a policy perspective. For instance, considering how city size affects environmental concerns or road congestion costs is important for designing urban policies that improve welfare.

There have been only a few early independent attempts to evaluate agglomeration costs and they are on developing countries only (Thomas, 1980; Richardson, 1987; Henderson, 2002). Recently, housing and land prices have started to be investigated more systematically, although articles usually rely for their analyses on datasets that are not comprehensive. There are a few rare exceptions such as Davis and Heathcote (2007) and Davis and Palumbo (2008) on the whole US, or Combes et al. (2012a) on the determinants of land prices in French urban areas. This last paper estimates the elasticity of land prices with respect to city population, from which the elasticity of urban costs is recovered. Its magnitude is found to be similar to the one of the elasticity of agglomeration gains on productivity. Albouy and Ehrlich (2013) replicate the approach to investigate the determinants of land prices in US metropolitan areas. Finally, some papers have tried to exploit natural or controlled experiments such as Rossi-Hansberg et al. (2010) who use residential urban revitalization programs implemented in Richmond, Virginia, to evaluate the effect of housing externalities on land value.

Housing is not the only good whose price varies across locations but little is known for other types of goods. Using barcode data on purchase transactions, Handbury and Weinstein (2014) and Handbury (2013) assess how prices of grocery products vary with city size. Handbury and Weinstein (2014) find that raw price indices slightly increase with city size, and this would constitute an additional source of agglomeration costs for households. However, this result is obtained before correcting prices for quality differences across varieties and before taking into account effects related to preferences for diversity that are present when considering CES utility functions. Once these are taken into account, price indices decrease with city size. This is the typical agglomeration gain that can be found in economic geography models with mobile workers *à la* Krugman (1991a). The price index decrease is due mostly to a much larger number of available varieties in larger cities but also to a higher quality of varieties sold there. Handbury (2013) allows preferences to differ between rich and poor households, and obtains the further result that the price index decreases with city size only for rich households but increase for poor ones. Clearly, investigating further these types of agglomeration effects is high on the agenda.

Last, since there is evidence that gains and costs from agglomeration as well location choices differ across types of workers, there is a need to consistently reintroduce space in welfare analyses when one wishes to assess individual or household inequalities. Moretti (2013) shows that real wage disparities between skilled and unskilled workers have increased less over the last 30 years than what nominal wage disparities would suggest, once the increase in the propensity of skilled

workers to live in larger cities compared to unskilled workers has been taken into account. Indeed, the increase in the difference in housing cost between skilled and unskilled workers represents up to 30% of the increase in the difference in nominal wages. Albouy et al. (2013) show that Canadian cities with the highest real wage differ for English- and French-speakers.

However, this type of real wage computation does not consider differences in amenity endowments across cities and possible differences in the valuation of amenities across worker groups. As workers are mobile, differences in real wages across locations should reflect to some extent differences in amenity value (see Roback, 1982). Albouy et al. (2013) show that indeed the real wage they compute for Canadian cities is slightly correlated with arts and climate city ratings. For the US, Albouy (2008) and Albouy (2009) find that the most valuable cities have coastal proximity, sunshine, and mild seasons. These findings are in line with those of Desmet and Rossi-Hansberg (2013) who use a slightly more general model calibrated on US data to assess the welfare impact of eliminating differences in amenities or frictions (within-city commuting time, local taxes, government expenditure) between cities. Diamond (2013) takes into account workers' heterogeneity and shows that the increased skill sorting in the US is partly due to the endogenous increase in amenities within higher skill cities.

Some recent theoretical contributions such as Behrens et al. (2014), Eeckhout et al. (2014), and Behrens and Robert-Nicoud (2014), suggest that sorting and disparities are worth studying simultaneously within and between cities. Glaeser et al. (2009) and Combes et al. (2012c) show that indeed larger cities present larger dispersions of wages and skills respectively in the US and France. Baum-Snow and Pavan (2013) further document the emergence of both within and between city inequalities in wages and skills in the US. A full empirical welfare assessment of both within and between city disparities considering agglomeration costs and benefits, heterogeneous workers that are imperfectly mobile, and amenity data in addition to productivity measures and land and housing prices, is a challenge for future research.

References

Abel, Jaison R., Dey, Ishita, and Gabe, Todd M. Productivity and the density of human capital. *Journal of Regional Science*, 52(4):562–586, 2012.

Abowd, John M., Kramarz, Francis, and Margolis, David N. High wage workers and high wage firms. *Econometrica*, 67(2):251–333, 1999.

Addario, Sabrina Di and Vuri, Daniela. Entrepreneurship and market size. The case of young college graduates in Italy. *Labour Economics*, 17(5):848–858, 2010.

Ahlfeldt, Gabriel, Redding, Stephen, Sturm, Daniel, and Wolf, Nikolaus. The economics of density: Evidence from the Berlin Wall. CEP Discussion Papers 1154, 2012.

Albouy, David. Are big cities really bad places to live? Improving quality-of-life estimates across cities. Working Paper 14472, National Bureau of Economic Research, 2008.

Albouy, David. What are cities worth? Land rents, local productivity, and the capitalization of amenity values. Working Paper 14981 Revised 2014, National Bureau of Economic Research, 2009.

Albouy, David and Ehrlich, Gabriel. The distribution of urban land values: Evidence from market transactions. Mimeograph, University of Illinois, 2013.

Albouy, David, Leibovici, Fernando, and Warman, Casey. Quality of life, firm productivity, and the value of amenities across Canadian cities. *Canadian Journal of Economics*, 46(2):379–411, 2013.

Amiti, Mary and Cameron, Lisa. Economic geography and wages. *Review of Economics and Statistics*, 89(1):15–29, 2007.

Ananat, Elizabeth, Fu, Shihe, and Ross, Stephen L. Race-specific agglomeration economies: Social distance and the black-white wage gap. Working Paper 18933, National Bureau of Economic Research, 2013.

Andersson, Fredrik, Burgess, Simon, and Lane, Julia I. Cities, matching and the productivity gains of agglomeration. *Journal of Urban Economics*, 61(1):112–128, 2007.

Andersson, Martin, Klaesson, Johan, and Larsson, Johan P. The sources of the urban wage premium by worker skills: Spatial sorting or agglomeration economies? *Papers in Regional Science*, page forthcoming, 2014.

Andini, Monica, de Blasio, Guido, Duranton, Gilles, and Strange, William. Marshallian labour market pooling: Evidence from Italy. *Regional Science and Urban Economics*, 43(6):1008–1022, 2013.

Arauzo-Carod, Josep-Maria and Viladecans-Marsal, Elisabet. Industrial location at the intra-metropolitan level: The role of agglomeration economies. *Regional Studies*, 43(4):545–558, 2009.

Arellano, Manuel and Bond, Steve. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies*, 58(2):277–297, 1991.

Arzaghi, Mohammad and Henderson, J. Vernon. Networking off Madison Avenue. *Review of Economic Studies*, 75(4):1011–1038, 2008.

Au, Chun-Chung and Henderson, J. Vernon. How migration restrictions limit agglomeration and productivity in China. *Journal of Development Economics*, 80(2):350–388, 2006a.

Au, Chun-Chung and Henderson, Vernon. Are Chinese cities too small? *Review of Economic Studies*, 73(3):549–576, 2006b.

Bacolod, Marigee, Blum, Bernardo S., and Strange, William C. Urban interactions: Soft skills versus specialization. *Journal of Economic Geography*, 9(2):227–262, 2009a.

Bacolod, Marigee, Blum, Bernardo S., and Strange, William C. Skills in the city. *Journal of Urban Economics*, 65(2):136 – 153, 2009b.

Bacolod, Marigee, Blum, Bernardo S., and Strange, William C. Elements of skills: Traits, intelligences, education, and agglomeration. *Journal of Regional Science*, 50(1):245–280, 2010.

Bai, Jushan. Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279, 2009.

Baldwin, John R., Brown, W. Mark, and Rigby, David L. Agglomeration economies: Microdata panel estimates from Canadian manufacturing. *Journal of Regional Science*, 50(5):915–934, 2010.

Barrios, Salvador, Görg, Holger, and Strobl, Eric. Multinationals' location choice, agglomeration economies, and public incentives. *International Regional Science Review*, 29(1):81–107, 2006.

Basile, Roberto. Acquisition versus greenfield investment: The location of foreign manufacturers in Italy. *Regional Science and Urban Economics*, 34(1):3–25, 2004.

Basile, Roberto, Castellani, Davide, and Zanfei, Antonello. Location choices of multinational firms in Europe: The role of EU cohesion policy. *Journal of International Economics*, 74(2):328–340, 2008.

Baum-Snow, Nathaniel and Ferreira, Fernando. Causal inference in urban economics. In Duranton, Gilles, Henderson, Vernon, and Strange, Will, editors, *Handbook of Urban and Regional Economics*, volume 5A. North-Holland, Amsterdam, 2015.

Baum-Snow, Nathaniel and Pavan, Ronni. Understanding the city size wage gap. *Review of Economic Studies*, 79(1):88–127, 2012.

Baum-Snow, Nathaniel and Pavan, Ronni. Inequality and city size. *Review of Economics and Statistics*, 93(5):1535–1548, 2013.

Beaudry, Paul, Green, David A., and Sand, Ben. Spatial equilibrium with unemployment and wage bargaining: Theory and estimation. *Journal of Urban Economics*, 79(C):2–19, 2014.

Behrens, Kristian and Robert-Nicoud, Frédéric. Survival of the fittest in cities: Urbanisation and inequality. *Economic Journal*, forthcoming, 2014.

Behrens, Kristian, Duranton, Gilles, and Robert-Nicoud, Frédéric. Productive cities: Sorting, selection, and agglomeration. *Journal of Political Economy*, 122(3):507–553, 2014.

Bleakley, Hoyt and Lin, Jeffrey. Thick-market effects and churning in the labor market: Evidence from US cities. *Journal of Urban Economics*, 72(2):87103, 2012.

Blien, Uwe and Suedekum, Jens. Local economic structure and industry development in Germany, 1993-2001. *Economics Bulletin*, 17(15):1–8, 2005.

Blien, Uwe, Suedekum, Jens, and Wolf, Katja. Productivity and the density of economic activity. *Labour Economics*, 13(4):445–458, 2006.

Bosker, Maarten, Brakman, Steven, Garretsen, Harry, and Schramm, Marc. Adding geography to the new economic geography: Bridging the gap between theory and empirics. *Journal of Economic Geography*, 10(6):793–823, 2010.

Brakman, Steven, Garretsen, Harry, and Schramm, Marc. The spatial distribution of wages: Estimating the Helpman-Hanson model for Germany. *Journal of Regional Science*, 44(3):437–466, 2004.

Brakman, Steven, Garretsen, Harry, and Schramm, Marc. Putting new economic geography to the test: Free-ness of trade and agglomeration in the {EU} regions. *Regional Science and Urban Economics*, 36(5):613–635, 2006.

Brakman, Steven, Garretsen, Harry, and Van Marrewijk, Charles. Economic geography within and between European nations: The role of market potential and density across space and time. *Journal of Regional Science*, 49(4):777–800, 2009.

Breinlich, Holger. The spatial income structure in the European Union - What role for economic geography? *Journal of Economic Geography*, 6(5):593–617, 2006.

Briant, Anthony, Combes, Pierre-Philippe, and Lafourcade, Miren. Does the size and shape of geographical units jeopardize economic geography estimations? *Journal of Urban Economics*, 67(3):287–302, 2010.

Brülhart, Marius and Mathys, Nicole A. Sectoral agglomeration economies in a panel of European regions. *Regional Science and Urban Economics*, 38(4):348 – 362, 2008.

Brunello, Giorgio and Gambarotto, Francesca. Do spatial agglomeration and local labor market competition affect employer-provided training? evidence from the UK. *Regional Science and Urban Economics*, 37(1):1–21, 2007.

Brunello, Giorgio and Paola, Maria De. Training and economic density: Some evidence form italian provinces. *Labour Economics*, 15(1):118–140, 2008.

Buchanan, James M. An economic theory of clubs. *Economica*, 32(125):1–14, 1965.

Carlino, Gerald and Kerr, William. Agglomeration and innovation. In Duranton, Gilles, Henderson, Vernon, and Strange, Will, editors, *Handbook of Urban and Regional Economics*, volume 5A. North-Holland, Amsterdam, 2015.

Carlsen, Fredrik, Rattsø, Jørn, and Stokke, Hildegunn E. Education, experience and dynamic urban wage premium. Department of Economics Working Paper 142013, Norwegian University of Science and Technology, 2013.

Carlton, Dennis. The location and employment choices of new firms: An econometric model with discrete and continuous endogenous variables. *Review of Economics and Statistics*, 65(3): 440–449, 1983.

Chauvin, Jean-Pierre, Glaeser, Edward, and Tobio, Kristina. Urban economics in the US and India. Work in progress, Harvard University, 2014.

Chinitz, Benjamin. Contrasts in agglomeration: New-York and Pittsburgh. *American Economic Review*, 51(2):279–289, 1961.

Ciccone, Antonio. Agglomeration effects in Europe. *European Economic Review*, 46(2):213–227, 2002.

Ciccone, Antonio and Hall, Robert E. Productivity and the density of economic activity. *American Economic Review*, 86(1):54–70, 1996.

Ciccone, Antonio and Peri, Giovanni. Identifying human capital externalities: Theory with an application to US cities. *Review of Economic Studies*, 73(2):381–412, 2006.

Cieślik, Andrzej. Regional characteristics and the location of foreign firms within poland. *Applied Economics*, 37(8):863–874, 2005.

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

Combes, Pierre-Philippe. Economic structure and local growth: France, 1984–1993. *Journal of Urban Economics*, 47(3):329–355, 2000.

Combes, Pierre-Philippe. The empirics of economic geography: How to draw policy implications? *Review of World Economics*, 147(3):567–592, 2011.

Combes, Pierre-Philippe and Duranton, Gilles. Labour pooling, labour poaching, and spatial clustering. *Regional Science and Urban Economics*, 36(1):1–28, 2006.

Combes, Pierre-Philippe and Lafourcade, Miren. Transport costs: Measures, determinants, and regional policy implications for France. *Journal of Economic Geography*, 5(3):319–349, 2005.

Combes, Pierre-Philippe and Lafourcade, Miren. Competition, market access and economic geography: Structural estimation and predictions for France. *Regional Science and Urban Economics*, 41:508–524, 2011.

Combes, Pierre-Philippe, Magnac, Thierry, and Robin, Jean-Marc. The dynamics of local employment in France. *Journal of Urban Economics*, 56(2):217–243, 2004.

Combes, Pierre-Philippe, Duranton, Gilles, and Gobillon, Laurent. Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, 63(2):723–742, 2008a.

Combes, Pierre-Philippe, Mayer, Thierry, and Thisse, Jacques-François. *Economic Geography: The integration of Regions and Nations*. Princeton University Press, New Jersey, 2008b.

Combes, Pierre-Philippe, Duranton, Gilles, Gobillon, Laurent, and Roux, Sébastien. Estimating agglomeration effects with history, geology, and worker fixed-effects. In Glaeser, Edward L., editor, *Agglomeration Economics*, pages 15–65. Chicago University Press, Chicago, IL, 2010.

Combes, Pierre-Philippe, Duranton, Gilles, and Gobillon, Laurent. The identification of agglomeration economies. *Journal of Economic Geography*, 11(2):253–266, 2011.

Combes, Pierre-Philippe, Duranton, Gilles, and Gobillon, Laurent. The costs of agglomeration: Land prices in French cities. Discussion Paper 9240, Centre for Economic Policy Research, 2012a.

Combes, Pierre-Philippe, Duranton, Gilles, Gobillon, Laurent, Puga, Diego, and Roux, Sébastien. The productivity advantages of large markets: Distinguishing agglomeration from firm selection. *Econometrica*, 80(6):2543–2594, 2012b.

Combes, Pierre-Philippe, Duranton, Gilles, Gobillon, Laurent, and Roux, Sébastien. Sorting and local wage and skill distributions in France. *Regional Science and Urban Economics*, 42(6): 913–930, 2012c.

Combes, Pierre-Philippe, Démurger, Sylvie, and Li, Shi. Urbanisation and migration externalities in China. Discussion Paper 9352, Centre for Economic Policy Research, 2013.

Corrado, Luisa and Fingleton, Bernard. Where is the economics in spatial econometrics? *Journal of Regional Science*, 52(2):210–239, 2012.

Crozet, Matthieu, Mayer, Thierry, and Mucchielli, Jean-Louis. How do firms agglomerate? A study of FDI in France. *Regional Science and Urban Economics*, 34(1):27 – 54, 2004.

Davis, Morris A. and Heathcote, Jonathan. The price and quantity of residential land in the United States. *Journal of Monetary Economics*, 54(8):2595–2620, 2007.

Davis, Morris A. and Palumbo, Michael G. The price of residential land in large US cities. *Journal of Urban Economics*, 63(1):352–384, 2008.

D'Costa, Sabine and Overman, Henry. The urban wage growth premium: Sorting or learning? *Regional Science and Urban Economics*, forthcoming, 2014.

De La Roca, Jorge and Puga, Diego. Learning by working in big cities. Discussion Paper 9243, Centre for Economic Policy Research, 2012.

Delgado, Mercedes, Porter, Michael E., and Stern, Scott. Clusters and entrepreneurship. *Journal of Economic Geography*, 10(4):495–518, 2010.

Desmet, Klaus and Fafchamps, Marcel. Changes in the spatial concentration of employment across US counties: A sectoral analysis 1972-2000. *Journal of Economic Geography*, 5(3):261–284, 2005.

Desmet, Klaus and Rossi-Hansberg, Esteban. Urban accounting and welfare. *American Economic Review*, 103(6):2296–2327, 2013.

Devereux, Michael P., Griffith, Rachel, and Simpson, Helen. Firm location decisions, regional grants and agglomeration externalities. *Journal of Public Economics*, 91(3-4):413 – 435, 2007.

Di Addario, Sabrina. Job search in thick markets. *Journal of Urban Economics*, 69(3):303–318, 2011.

Di Addario, Sabrina and Patacchini, Eleonora. Wages and the city. Evidence from Italy. *Labour Economics*, 15(5):1040–1061, 2008.

Di Giacinto, Valter, Gomellini, Matteo, Micucci, Giacinto, and Pagnini, Marcello. Mapping local productivity advantages in italy: Industrial districts, cities or both? *Journal of Economic Geography*, (2):365394, 2014.

Diamond, Rebecca. The determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000. Mimeograph, Stanford University, 2013.

Drucker, Joshua and Feser, Edward. Regional industrial structure and agglomeration economies: An analysis of productivity in three manufacturing industries. *Regional Science and Urban Economics*, 42(12):1–14, 2012.

Dumais, Guy, Ellison, Glenn, and Glaeser, Edward L. Geographic concentration as a dynamic process. Working Paper 6270, National Bureau of Economic Research, 1997.

Dumais, Guy, Ellison, Glenn, and Glaeser, Edward L. Geographic concentration as a dynamic process. *Review of Economics and Statistics*, 84(2):193–204, 2002.

Duranton, Gilles. Agglomeration effects in Colombia. Mimeograph, Wharton University, 2014.

Duranton, Gilles and Jayet, Hubert. Is the division of labour limited by the extent of the market? Evidence from French cities. *Journal of Urban Economics*, 69(1):56–71, 2011.

Duranton, Gilles and Monastiriotis, Vassilis. Mind the gaps: The evolution of regional earnings inequalities in the UK 1982-1997. *Journal of Regional Science*, 42(2):219–256, 2002.

Duranton, Gilles and Overman, Henry G. Testing for localization using micro-geographic data. *Review of Economic Studies*, 72(4):1077–1106, 2005.

Duranton, Gilles and Puga, Diego. Nursery cities: Urban diversity, process innovation, and the life cycle of products. *American Economic Review*, 91(5):1454–1477, 2001.

Duranton, Gilles and Puga, Diego. Micro-foundations of urban agglomeration economies. In Henderson, J. Vernon and Thisse, Jacques-François, editors, *Handbook of Regional and Urban Economics*, volume 4, pages 2063–2117. North-Holland, Amsterdam, 2004.

Eeckhout, Jan, Pinheiro, Roberto, and Schmidheiny, Kurt. Spatial sorting. *Journal of Political Economy*, 122(3):554–620, 2014.

Ehrl, Philipp. Agglomeration economies with consistent productivity estimates. *Regional Science and Urban Economics*, 43(5):751–763, 2013.

Ellison, Glenn and Glaeser, Edward L. Geographic concentration in US manufacturing industries: A dartboard approach. *Journal of Political Economy*, 105(5):889–927, 1997.

Ellison, Glenn and Glaeser, Edward L. The geographic concentration of industry: Does natural advantage explain agglomeration? *American Economic Review Papers and Proceedings*, 89(2): 311–316, 1999.

Ellison, Glenn, Glaeser, Edward L., and Kerr, William R. What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review*, 100(3):1195–1213, 2010.

Elvery, Joel A. City size and skill intensity. *Regional Science and Urban Economics*, 40(6):367–379, 2010.

Faberman, Jason and Freedman, Matthew. The urban density premium across establishments. Working Paper 2013-01, Federal Reserve Bank of Chicago, 2013.

Fallah, Belal N., Partridge, Mark D., and Olfert, M. Rose. New economic geography and US metropolitan wage inequality. *Journal of Economic Geography*, 46(5):865–895, 2011.

Fally, Thibault, Paillacar, Rodrigo, and Terra, Cristina. Economic geography and wages in Brazil: Evidence from micro-data. *Journal of Development Economics*, 91(1):155–168, 2010.

Figueiredo, Octávio, Guimarães, Paulo, and Woodward, Douglas. Home-field advantage: Location decisions of Portuguese entrepreneurs. *Journal of Urban Economics*, 52(2):341–361, 2002.

Figueiredo, Octávio, Guimarães, Paulo, and Woodward, Douglas. Firmworker matching in industrial clusters. *Journal of Economic Geography*, 14(1):1–19, 2014.

Foster, Neil and Stehrer, Robert. Sectoral productivity, density and agglomeration in the Wider Europe. *Spatial Economic Analysis*, 4(4):427–446, 2009.

Freedman, Matthew. Job hopping, earnings dynamics, and industrial agglomeration in the software publishing industry. *Journal of Urban Economics*, 64(3):590–600, 2008.

Fu, Shihe. Smart café cities: Testing human capital externalities in the Boston metropolitan area. *Journal of Urban Economics*, 61(1):86–111, 2007.

Fu, Shihe and Ross, Stephen L. Wage premia in employment clusters: How important is worker heterogeneity? *Journal of Labor Economics*, 31(2):271–304, 2013.

Fuchs, Michaela. The determinants of local employment dynamics in Western Germany. *Empirical Economics*, 40(1):177–203, 2011.

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

Ghani, Ejaz, Kerr, William, and O'Connell, Stephen. Local industrial structures and female entrepreneurship in India. *Journal of Economic Geography*, 13(6):929–964, 2013.

Ghani, Ejaz, Kerr, William, and O'Connell, Stephen. Determinants of entrepreneurship in India. *Regional Studies*, 48(6):1071–1089, 2014.

Gibbons, Stephen and Overman, Henry G. Mostly pointless spatial econometrics. *Journal of Regional Science*, 52(2):172–191, 2012.

Gibbons, Stephen, Overman, Henry G., and Resende, Guilherme. Real earnings disparities in Britain. Discussion Paper 65, Spatial Economic Research Center, 2011.

Glaeser, Edward L. and Kerr, William R. Local industrial conditions and entrepreneurship: How much of the spatial distribution can we explain? *Journal of Economics an dManagement Strategy*, 18(3):623–663, 2009.

Glaeser, Edward L. and Maré, David C. Cities and skills. *Journal of Labor Economics*, 19(2): 316–342, 2001.

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

Glaeser, Edward L., Resseger, Matt, and Tobio, Kristina. Inequality in cities. *Journal of Regional Science*, 49(4):617–646, 2009.

Glaeser, Edward L., Kerr, William R., and Ponzetto, Giacomo A.M. Clusters of entrepreneurship. *Journal of Urban Economics*, 67(1):150–168, 2010a.

Glaeser, Edward L., Rosenthal, Stuart S., and Strange, William C. Urban economics and entrepreneurship. *Journal of Urban Economics*, 67(1):1–14, 2010b.

Gould, Eric. Cities, workers, and wages: A structural analysis of the urban wage premium. *Review of Economic Studies*, 74:477–506, 2007.

Graham, Daniel J. Variable returns to agglomeration and the effect of road traffic congestion. *Journal of Urban Economics*, 62(1):103–120, 2007.

Graham, Daniel J. Identifying urbanisation and localisation externalities in manufacturing and service industries. *Papers in Regional Science*, 88(1):63–84, 2009.

Graham, Daniel J., Melo, Patricia S., Jiwattanakulpaisarn, Piyapong, and Noland, Robert B. Testing for causality between productivity and agglomeration economies. *Journal of Regional Science*, 50(5):935–951, 2010.

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

Groot, Stefan P.T. and de Groot, Henri L.F. Estimating the skill bias in agglomeration externalities and social returns to education: Evidence from dutch matched worker-firm micro-data. Discussion Paper 2014-088, Tinbergen Institute, 2014.

Groot, Stefan P.T., de Groot, Henri L.F., and Smit, Martijn J. Reginal wage differences in the Netherlands: Micro evidence on agglomeration externalities. *Journal of Regional Science*, 54(3): 503–523, 2014. ISSN 1467-9787.

Guimaraes, Paulo, Figueiredo, Octávio, and Woodward, Douglas. Agglomeration and the location of foreign direct investment in Portugal. *Journal of Urban Economics*, 47(1):115–135, 2000.

Handbury, Jessie. Are poor cities cheap for everyone? Non-homotheticity and the cost of living across US cities. Mimeograph, Wharton University, 2013.

Handbury, Jessie and Weinstein, David. Goods prices and availability in cities. *Quarterly Journal of Economics*, 2014. forthcoming.

Hanson, Gordon H. Increasing returns, trade, and the regional structure of wages. *Economic Journal*, 107(440):113–133, 1997.

Hanson, Gordon H. Market potential, increasing returns, and geographic concentration. *Journal of International Economics*, 67(1):1–24, 2005.

Harada, Nobuyuki. Potential entrepreneurship in Japan. *Small Business Economics*, 25:293–304, 2005.

Harris, Chauncy. The market as a factor in the localization of industry in the United States. *Annals of the Association of American Geographers*, 44(4):315–348, 1954.

Head, Keith and Mayer, Thierry. Market potential and the location of Japanese investment in the European Union. *Review of Economics and Statistics*, 86(4):959–972, 2004.

Head, Keith and Mayer, Thierry. Regional wage and employment responses to market potential in the EU. *Regional Science and Urban Economics*, 36(5):573–595, 2006.

Head, Keith, Ries, John C., and Swenson, Deborah L. Attracting foreign manufacturing: Investment promotion and agglomeration. *Regional Science and Urban Economics*, 29(2):197–218, 1999.

Heckman, James and Singer, Burton. A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, 2(52):271–320, 1984.

Henderson, J. Vernon. Externalities and industrial development. *Journal of Urban Economics*, 42 (3):449–470, 1997.

Henderson, J. Vernon. Marshall's economies. *Journal of Urban Economics*, 53(1):1–28, 2003.

Henderson, J. Vernon, Kuncoro, Ari, and Turner, Matt. Industrial development in cities. *Journal of Political Economy*, 103(5):1067–1090, 1995.

Henderson, Vernon. Urban primacy, external costs, and the quality of life. *Resource and Energy Economics*, 24(1):95–106, 2002.

Henderson, Vernon, Lee, Todd, and Lee, Yung Joon. Scale externalities in Korea. *Journal of Urban Economics*, 49:479–504, 2001.

Hering, Laura and Poncet, Sandra. Market access and individual wages: Evidence from China. *Review of Economics and Statistics,*, 92:145–159, 2010.

Hilber, Christian A. L. and Voicu, Ioan. Agglomeration economies and the location of foreign direct investment: Empirical evidence from Romania. *Regional Studies*, 44(3):355–371, 2010.

Holl, Adelheid. Market potential and firm-level productivity in Spain. *Journal of Economic Geography*, 12(6):1191–1215, 2012.

Holmes, Thomas and Sieg, Holger. Structural estimation in urban economics. In Duranton, Gilles, Henderson, Vernon, and Strange, Will, editors, *Handbook of Urban and Regional Economics*, volume 5A. North-Holland, Amsterdam, 2015.

Holmes, Thomas J. The diffusion of Wal-Mart and economies of density. *Econometrica*, 79(1): 253–302, 2011.

Holmes, Thomas J. and Lee, Sanghoon. Economies of density versus natural advantage: Crop choice on the Back Forty. *Review of Economics and Statistics*, 94(1):1–19, 2012.

Imbens, Guido and Angrist, Joshua. Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475, 1994.

Jacobs, Jane. *The Economy of Cities*. Random House, New York, 1969.

Jofre-Montseny, Jordi, Marin-Lopez, Raquel, and Viladecans-Marsal, Elisabet. The mechanisms of agglomeration: Evidence from the effect of inter-industry relations on the location of new firms. *Journal of Urban Economics*, 70(2-3):61–74, 2011.

Jofre-Montseny, Jordi, Marín-López, Raquel, and Viladecans-Marsal, Elisabet. The determinants of localization and urbanization economies: Evidence from the location of new firms in Spain. *Journal of Regional Science*, 54(2):313–337, 2014.

Kerr, William and Kominers, Scott Duke. Agglomerative forces and cluster shapes. *Review of Economics and Statistics*, forthcoming, 2014.

Kim, Sukkoo. Expansion of markets and the geographic distribution of economic activities: The trends in US regional manufacturing structure, 1860-1987. *Quarterly Journal of Economics*, 110 (4):881–908, 1995.

Kok, Suzanne. Town and city jobs: Your job is different in another location. *Regional Science and Urban Economics*, forthcoming, 2015.

Kolko, Jed. Urbanization, agglomeration, and coagglomeration of service industries. In Glaeser, Edward L., editor, *The Economics of Agglomeration*, pages 151–180. National Bureau of Economic Research, Cambridge (MASS.), 2010.

Krashinsky, Harry. Urban agglomeration, wages and selection: Evidence from samples of siblings. *Labour Economics*, 18(1):79–92, 2011.

Krugman, Paul R. Increasing returns and economic geography. *Journal of Political Economy*, 99 (3):484–499, 1991a.

Krugman, Paul R. *Geography and Trade*. MIT Press, Cambridge, MA, 1991b.

Lall, Somik V., Shalizi, Zmarak, and Deichmann, Uwe. Agglomeration economies and productivity in Indian industry. *Journal of Development Economics*, 73(3):643–673, 2004.

Lee, Bun Song, Jang, Soomyung, and Hong, Sung Hyo. Marshalls scale economies and Jacobs externality in Korea: The role of age, size and the legal form of organisation of establishments. *Urban Studies*, 47(14):3131–3156, 2010.

Lee, Sanghoon. Ability sorting and consumer city. *Journal of Urban Economics*, 68(1):20–33, 2010.

Lehmer, Florian and Möller, Joachim. Interrelations between the urban wage premium and firm-size wage differentials: a microdata cohort analysis for germany. *The Annals of Regional Science*, 45(1):31–53, 2010.

Lesage, James and Pace, Robert Kelley. *Introduction to spatial econometrics*. CRC Press, New York, 2009.

Levinsohn, James and Petrin, Amil. Estimating production functions using inputs to control for unobservables. *Review of Economic Studies*, 70(2):317–342, 2003.

Lindley, Joanne and Machin, Stephen. Spatial changes in labour market inequality. *Journal of Urban Economics*, 79(0):121–138, 2014. Spatial Dimensions of Labor Markets.

Lopez, Ricardo A. and Suedekum, Jens. Vertical industry relations, spillovers, and productivity: Evidence from Chilean plants. *Journal of Regional Science*, 49(4):721–747, 2009.

Lucas, Robert E. Jr. On the mechanics of economic development. *Journal of Monetary Economics*, 22(1):3–42, 1988.

Mameli, Francesca, Faggian, Alessandra, and McCann, Philip. Employment growth in Italian local labour systems: Issues of model specification and sectoral aggregation. *Spatial Economic Analysis*, 3(3):343–360, 2008.

Marrocu, Emanuela, Paci, Raffaele, and Usai, Stefano. Productivity growth in the Old and New Europe: The role of agglomeration externalities. *Journal of Regional Science*, 53(3):418–442, 2013.

Marshall, Alfred. *Principles of Economics*. Macmillan, London, 1890.

Martin, Philippe, Mayer, Thierry, and Mayneris, Florian. Spatial concentration and plant-level productivity in France. *Journal of Urban Economics*, 69(2):182–195, 2011.

Matano, Alessia and Naticchioni, Paolo. Wage distribution and the spatial sorting of workers. *Journal of Economic Geography*, 12(2):379–408, 2012.

McFadden, Dennis. Conditional logit analysis of qualitative choice behavior. In Zarembka, Paul, editor, *Frontier in Econometrics*, pages 105–142. Academic Press, New York, 1974.

McMillen, Daniel P. Perspectives on spatial econometrics: Linear smoothing with structured models. *Journal of Regional Science*, 52(2):192–209, 2012.

Melitz, Marc J. and Ottaviano, Gianmarco I. P. Market size, trade, and productivity. *Review of Economic Studies*, 75(1):295–316, 2008.

Melo, Patricia C., Graham, Daniel J., and Noland, Robert B. A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics*, 39(3):332–342, 2009.

Micucci, Giacinto and Giacinto, Valter di. The producer service sector in Italy: Long-term growth and its local determinants. *Spatial Economic Analysis*, 4(4):391–425, 2009.

Mion, Giordano. Spatial externalities and empirical analysis: The case of Italy. *Journal of Urban Economics*, 56(1):97–118, 2004.

Mion, Giordano and Naticchioni, Paolo. The spatial sorting and matching of skills and firms. *Canadian Journal of Economics*, 42(1):28–55, 2009.

Moretti, Enrico. Workers' education, spillovers, and productivity: Evidence from plant-level production functions. *American Economic Review*, 94(3):656–690, 2004a.

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

Moretti, Enrico. Real wage inequality. *American Economic Journal: Applied Economics*, 5(1): 65–103, 2013.

Morikawa, Masayuki. Economies of density and productivity in service industries: An analysis of personal service industries based on establishment-level data. *Review of Economics and Statistics*, 93(1):179–192, 2011.

Moulton, Brent R. An illustration of the pitfall in estimating the effects of aggregate variables on micro units. *Review of Economics and Statistics*, 72(2):334–338, 1990.

Muehlemann, Samuel and Wolter, Stefan C. Firm-sponsored training and poaching externalities in regional labor markets. *Regional Science and Urban Economics*, 41(6):560–570, 2011.

Neumark, David and Simpson, Helen. Place-based policies. In Duranton, Gilles, Henderson, Vernon, and Strange, Will, editors, *Handbook of Urban and Regional Economics*, volume 5A. North-Holland, Amsterdam, 2015.

Olley, G. Steven and Pakes, Ariel. The dynamics of productivity in the telecommunication equipment industry. *Econometrica*, 64(6):1263–1297, 1996.

Overman, Henry G. and Puga, Diego. Labor pooling as a source of agglomeration: An empirical investigation. In Glaeser, Edward L., editor, *The Economics of Agglomeration*, pages 133–150. National Bureau of Economic Research, Cambridge (MASS.), 2010.

Phimister, Euan. Urban effects on participation and wages: Are there gender differences? *Journal of Urban Economics*, 58(3):513–536, 2005.

Pusterla, Fazia and Resmini, Laura. Where do foreign firms locate in transition countries? An empirical investigation. *The Annals of Regional Science*, 41(4):835–856, 2007.

Redding, Stephen and Sturm, Daniel. The costs of remoteness: Evidence from German division and reunification. *American Economic Review*, 98(5):1766–1797, 2008.

Redding, Stephen and Venables, Anthony J. Economic geography and international inequality. *Journal of International Economics*, 62(1):63–82, 2004.

Rice, Patricia, Venables, Anthony J., and Patacchini, Eleonora. Spatial determinants of productivity: Analysis for the regions of Great Britain. *Regional Science and Urban Economics*, 36(6): 727–752, 2006.

Richardson, Harry W. The costs of urbanization: A four-country comparison. *Economic Development and Cultural Change*, 35(3):561–580, 1987.

Rigby, David L. and Essletzbichler, Jrgen. Agglomeration economies and productivity differences in US cities. *Journal of Economic Geography*, 2(4):407–432, 2002.

Roback, Jennifer. Wages, rents and the quality of life. *Journal of Political Economy*, 90(6): 1257–1278, 1982.

Rodríguez-Pose, Andrès and Tselios, Vassilis. Individual earnings and educational externalities in the european union. *Regional Studies*, 46(1):39–57, 2012.

Rosenthal, Stuart S. and Strange, William C. The determinants of agglomeration. *Journal of Urban Economics*, 50(2):191–229, 2001.

Rosenthal, Stuart S. and Strange, William C. Geography, industrial agglomeration, and agglomeration. *Review of Economics and Statistics*, 85(2):377–393, 2003.

Rosenthal, Stuart S. and Strange, William C. Evidence on the nature and sources of agglomeration economies. In Henderson, Vernon and Thisse, Jacques-François, editors, *Handbook of Regional and Urban Economics*, volume 4, pages 2119–2171. North-Holland, Amsterdam, 2004.

Rosenthal, Stuart S. and Strange, William C. The attenuation of human capital spillovers. *Journal of Urban Economics*, 64(2):373–389, 2008.

Rossi-Hansberg, Esteban, Sarte, Pierre-Daniel, and Owens III, Raymond. Housing externalities. *Journal of Political Economy*, 118(3):485–535, 2010.

Saito, Hisamitsu and Gopinath, Munisamy. Plants self-selection, agglomeration economies and regional productivity in Chile. *Journal of Economic Geography*, 9(4):539–558, 2009.

Sato, Yasuhiro, Tabuchi, Takatoshi, and Yamamoto, Kazuhiro. Market size and entrepreneurship. *Journal of Economic Geography*, 12(6):1139–1166, 2012.

Serafinelli, Michel. Good firms, worker flows and local productivity. Mimeograph, University of Toronto, 2014.

Simon, Curtis J. Industrial reallocation across US cities, 1977-1997. *Journal of Urban Economics*, 56(1):119–143, 2004.

Solé-Ollé, Albert and Viladecans-Marsal, Elisabet. Central cities as engines of metropolitan area growth. *Journal of Regional Science*, 44(2):321–350, 2004.

Spies, Julia. Network and border effects: Where do foreign multinationals locate in Germany? *Regional Science and Urban Economics*, 40(1):20 – 32, 2010.

Stock, James H. and Yogo, Motohiro. Testing for weak instruments in linear IV regression. In Andrews, Donald W.K. and Stock, James H., editors, *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, pages 80–108. Cambridge University Press, Cambridge, 2005.

Suedekum, Jens. Convergence of the skill composition across German regions. *Regional Science and Urban Economics*, 38(2):148–159, 2008.

Suedekum, Jens. Human capital externalities and growth of high- and low-skilled jobs. *Jahrbucher fur Nationalokonomie und Statistik*, 230(1):92–114, 2010.

Thomas, Vinod. Spatial differences in the cost of living. *Journal of Urban Economics*, 8(1): 108–122, 1980.

Van Soest, Daan P., Gerking, Shelby, and Van Oort, Frank G. Spatial impact of agglomeration externalities. *Journal of Regional Science*, 46(5):881–899, 2006.

Viladecans-Marsal, Elisabet. Agglomeration economies and industrial location: City-level evidence. *Journal of Economic Geography*, 5(4):565–582, 2004.

Wang, Zhi. Smart city: Learning effects and labor force entry. Mimeograph, Brown University, 2013.

Wheaton, William C. and Lewis, Mark J. Urban wages and labor market agglomeration. *Journal of Urban Economics*, 51(3):542–562, 2002.

Wheeler, Christopher. Local market scale and the pattern of job changes among young men. *Regional Science and Urban Economics*, 38(2):101118, 2008.

Wheeler, Christopher H. Cities and the growth of wages among young workers: Evidence from the NLSY. *Journal of Urban Economics*, 60(2):162 – 184, 2006.

Yankow, Jeffrey J. Why do cities pay more? An empirical examination of some competing theories of the urban wage premium. *Journal of Urban Economics*, 60(2):139 – 161, 2006.