Skills, tasks and skill-biased technological change in cities

Hug, Jessica

Skills, tasks and skill-biased technological change in cities

(Compétences, tâches et changement technologique dans les villes)

by

Jessica Hug

A thesis submitted to the
Geneva School of Economics and Management,
University of Geneva, Switzerland,
in fulfillment of the requirements for the degree of
PhD in Economics

Members of the thesis committee:
Prof. Sylvie Charlot, University Lyon 2 GATE Lyon St Etienne - UMR
Prof. Michele Pellizzari, Chair, University of Geneva
Prof. Frédéric Robert-Nicoud, Adviser, University of Geneva
Prof. Céline Carrère, Co-adviser, University of Geneva

Thesis No. 73
November 2019

La Faculté d’économie et de management, sur préavis du jury, a autorisé l’impression de la présente thèse, sans entendre, par-là, émettre aucune opinion sur les propositions qui s’y trouvent énoncées et qui n’engagent que la responsabilité de leur auteur.

Geneva, le 22 novembre 2019

Dean
Marcelo OLARREAGA

Impression d’après le manuscrit de l’auteur

Acknowledgements

I would like to thank Jonathan Dingel, Sylvie Charlot, Amit Khandelwal, Eduardo Morales, Tuan Nguyen, Michele Pellizzari, Esteban Rossi-Hansberg and participants of the 2017 NARSC conference, Vancouver (Urban Economic Association), EALE, St.Gallen, EEA, Lisbon, Summer School in Urban Economics, Paris, European Urban Economics Association Meeting, Copenhagen, Rhône Spatial Economics workshop, Geneva, Young Swiss Economics Meeting, Zurich, 2016 ERSA conference in Vienna and NARSC conference in Minneapolis for very helpful discussions. I also thank Henry Overman for hosting me and supervising me while visiting the London School of Economics and Political Science as well as my two co-supervisors Céline Carrère and Frédéric Robert-Nicoud for continuous support and guidance. The use of the data in this dissertation does not imply the endorsement of the data owner or the UK Data Service at the UK Data Archive in relation to the interpretation or analysis of the data. This work uses research datasets which may not exactly reproduce National Statistics aggregate.

Abstract

This dissertation develops a model investigating how computerization shapes the distribution of economic activities across space. Computer capital is a complement to complex tasks and these tasks generate and benefit from knowledge spillovers in large agglomerations. Thus, the complementarity between computer capital and city size is skill-biased where workers with a comparative advantage in those tasks disproportionately sort in large cities. Data from the United Kingdom and Germany support a complementarity between computer capital and city size. Indeed, an increase in city size is associated with a significant increase in the probability of different indicators of computer capital: using a computer at work, having computerized equipment as the most important work equipment, having an advanced use computer at work. The positive relationship holds even controlling for detailed workers and jobs characteristics. Consistent with housings costs being the dispersive force linked to computerization for low-skilled workers, the complementarity between computer use at work and city size is positively significant for them also. Finally, I develop a structural model where parameters are estimated empirically to quantify the impact of computerization on the increasing sorting of high-skilled workers in large cities.

Résumé

La présente dissertation doctorale développe un modèle pour étudier les effets de l'informatisation sur la distribution spatiale des activités économiques. Le capital informatique est complémentaire avec les tâches complexes et ce type de tâches bénéficie des externalités de connaissances présentes dans les grandes agglomérations. Par conséquent, la complémentarité entre le capital informatique et la taille de la ville est biaisée en faveur des travailleurs avec de hauts niveaux de formation et doués d'un avantage comparatif avec ces tâches. Ils sont donc surreprésentés dans les grandes agglomérations. Des données provenant du Royaume-Uni et de l'Allemagne démontre que le capital informatique est complémentaire avec la taille de la ville: lorsqu'elle augmente, la probabilité de travailler avec un ordinateur, d'avoir un équipement informatique comme équipement principal de travail, ou avoir un usage avancé de l'ordinateur augmente de façon significative. La relation positive persiste dans une régression qui contrôle pour des caractéristiques détaillées de l'individu et de son entreprise. La complémentarité entre l'usage de l'ordinateur au travail et la taille de la ville est présente également pour les travailleurs ayant fait des études plus courtes. Malgré tout, le coût du logement en augmentation suite à l'arrivée de travailleurs très formés rendrait les grandes villes moins attractives pour eux. Finalement, je développe un modèle structurel dans lequel les paramètres sont estimés empiriquement pour quantifier l'impact de l'informatisation de l'économie sur l'augmentation de la proportion de travailleurs très formés dans les grandes villes.

Contents

Introduction

This dissertation explores computerization as one driver of the choice of high-skilled workers to work and live in large agglomerations. Computer capital is a complement to complex tasks, which high-skilled workers have a comparative advantage with, and these tasks generate and benefit from knowledge spillovers in large cities. Thus, the complementarity between computer capital and city size is skill-biased where workers with a comparative advantage in those tasks disproportionately sort in large cities.

Understanding the sorting of high-skilled workers across city sizes is important because it is informative about real wage inequalities. As local living costs varies, nominal wage inequalities do not reflect inequalities in purchasing power across skill groups and regions. If computerization shapes the choice of high-skilled workers to live and work in large agglomerations, it will impact local living costs and real wage inequalities. Moreover, computerization would lead to an increasingly segregated population across agglomerations of different sizes within a country according to education level. As such, population with different educations have less opportunity to meet, go to school together or be confronted with the same local challenges. They could thus live in increasingly hermetic world, failing to understand each others. Last, if firms invest more in computer capital in large cities, differences in urbanization rates can explain differences in labor productivity related to differences in investment in computerized equipment.

I define a city (or interchangeably an agglomeration) as an integrated labor market based on commuting flows. As such a rural area can also be an agglomeration if most of the workforce live and work there. Next, I define a complex task as a non-routine task, that is not "sufficiently well understood that the task can be fully specified as a series of instructions to be executed by a machine" and requires "problem-solving, intuition, persuasion, and creativity" (Acemoglu and Autor (2011), Section 2.5), cognitive tasks. These tasks can be interactive or analytical and are also the ones which are most difficult to off-shore. I refer to all other tasks as simple.[1] Likewise, I refer to complex capital as capital which is used jointly with complex tasks to produce a complex output. It includes computer capital. Simple capital is used jointly with simple tasks to produce a simple output.

In Chapter 1, I formalize the complementarity between computer capital and city size in a spatial equilibrium model where workers with different skill levels chose in which agglomeration to work and live. As the price of computerized equipment decreases relative to simple equipment, firms increase their relative investment in computerized equipment, and especially so in large cities where its productivity is comparatively large. Because

[1]See Autor (2013) for a review of the "task approach" to the labor market.

high-skilled workers allocate more time to complex tasks, complementary with computerized equipment, they benefit more from the additional productivity in large cities due to the additional investment in computerized equipment there. Consequently, the proportion of high-skilled workers in large cities increases. Because of limited housing, large cities become less attractive to lower-skilled workers.

Next, I document the prediction of the model that states that firms allocate more computer capital in large cities relative to the rest of the country to a worker with given characteristics. In this exercise, I proxy an increase in computer capital with an increase in the probability to have an advanced use of computer at work, such as analyzing data. Indeed, it requires more computer power and as such more investment in computer capital compared to basic uses such as word processing. The main advantage of this proxy is that it allows me to capture a variation in computer capital for workers with a university degree where almost every worker uses a computer at work. I use the British Skill Survey and the German Working population Survey, which include detailed information on computer use at work. It also allows me to observe within occupation variation of tasks at work across city sizes. Moreover, it contains detailed workers' characteristics, such as overall score on leaving certificate or the type of school attended. The main caveat of my analysis is that I cannot exclude unobserved workers' characteristics as the driver of the complementarity between city size and computer capital.

Then, I expose potential mechanisms explaining the empirical complementarity between city size and computer capital. Beside complex tasks, I investigate whether labor capital complementarity in the presence of agglomeration economies, differences in the size of equipment or differences in infrastructures across cities can explain the positive relationship between computer capital and city size.

I finally observe in the British Skill Survey that individual wages are positively related to the triple interaction of city size, computer capital and skill. Although I cannot infer causality, it suggests that computerization is a valid candidate to partly explain the choice of high skilled workers to live in large cities.

In Chapter 2, I provide a framework to causally quantify the increasing proportion of high-skilled workers in large cities related to computerization. In this exercise, I do not go into the source of the complementarity between computer capital and city size. I extend the framework of Burstein et al. (2019) about workers' skill-equipment-occupation complementarities to include a spatial dimension with settlement types. The first purpose of the framework is to simulate the change in the share of high-skilled workers in large cities absent the decline in the relative price of computerized equipment. Its second purpose is to simulate wage inequality as a function of city size absent this decline.

The method used is one way to test assignment models as suggested by Acemoglu and Autor (2011) and addresses the drawbacks highlighted by DiNardo and Pischke (1997) when trying to capture the effect of computerization on wages. In the model, I express the change in the proportion of high-skilled workers in large cities as a function of (i) the change in the relative price of computerized equipment common for all workers and (ii) measures of workers specialization across equipment types in the first period available in the data. The latter captures the complementarity between equipment types, occupation and agglomeration types. The key hypothesis is thus that it is time invariant. Changes in local unobserved skill is captured by a residual term in the wage equation and is kept constant in the simulations.

Last, I illustrate how the model can be taken to the data to perform the simulations using data from Germany. I describe how to recover the complementarity between computer capital and city size through measures of workers specialization across equipment and agglomeration types. I also show how to retrieve the change in the relative price of computerized equipment based on the model. Finally, I show how to estimate the two key elasticities which govern the strength of the response of computerization on the skill sorting across city sizes. The first one is the response of wage to changes in equipment prices or productivities. The second one is the response of the share of workers with a university degree to changes in their local wages relative to workers without a university degree.

Chapter 1

City-biased technological change

1.1 Introduction

This paper introduces a new theoretical link between computerization and economic incentives to agglomerate in large cities for high-skilled workers. High-skilled workers have a comparative advantage performing complex tasks relative to simple ones and complex tasks are used jointly with computer capital in production. Moreover, complex tasks disproportionately benefit in large cities from learning externalities. As a result, computerization increases the attractivity of large cities for high-skilled workers. To illustrate the mechanism, consider an oncologist working at identifying the type of cancer she is confronted with. A computer program of medical imaging allowing her to visualize the atypical cells from every angle makes her more productive at the task. Nevertheless, the task is complex with part of the knowledge being tacit such as experience. As a result, the knowledge spillovers in the environment of large cities are particularly beneficial to perform it. The incentives to invest in the computer program are larger there and the strongest the comparative advantage of oncologist with computer capital through complex tasks relative to other workers, the larger their employment share in large cities compared to smaller ones.

I formalize this mechanism using a model with the following features. Fully mobile workers heterogeneous in their skill level choose a city which have exogenous productivities reflecting for example natural amenities. Workers choose their location by trading-off the consumption of a fully tradable final good and housing. Given the choice of the city, they choose the share of time to allocate to complex tasks and a representative firm chooses the quantity of computer and simple capital to assign to the workers in order to produce an intermediate good which is non-tradable. The productivity of the complex tasks - computer capital couple positively depends on the skill of the worker as well as on city size which reflects agglomeration economies and represents the endogenous part of the productivity of cities. Skill types are complement in production at the local level. This embodies the imperfect substitutability across skill groups at the local level supported by existing empirical evidence (See e.g. Ciccone and Peri (2006) for a discussion and an application within the US.). The approach provides a new way of ensuring the existence of a spatial equilibrium with many skill levels in all cities. Alternative specifications to model the spatial sorting of heterogeneous agents in a perfect mobility setting include Behrens et al. (2014), Davis and Dingel (2019), Davis and Dingel (2017), Eeckhout et al.

(2014) and Gaubert (2018).[1]

The equilibrium of the model displays the following properties: (i) *the population increases with the exogenous productivity of the site*, (ii) *the proportion of high skilled workers increases with city size*, (iii) *the representative firm allocates more computerized equipment to workers in large cities compared to smaller ones, for a given skill type* and (iv) *The proportion of high-skilled workers in large cities increases as the ratio of prices of simple to computer capital increases*. Indeed, as high-skilled workers allocate more time to complex tasks, they disproportionately benefit from the additional investment in computer capital in large cities. (i) and (ii) are widely documented[2], while (iii) and (iv) are novel. I take (iii) to the British Skill Survey and the German Working population Survey for the year 2006, where I proxy an increase in computer capital with an increase in the probability to have an advanced use of computer. Indeed, it requires more computer power and as such more investment in computer capital compared to basic uses.

I find that a German worker with mean characteristics has a probability of 0.24 instead of 0.20 to have an advanced computer use at work in large urban centers relative to the rest of the country. Likewise, a British worker with mean characteristics has a probability of 0.19 instead of 0.16 to have an advanced computer use at work in large urban centers relative to the rest of the country. All education categories are affected by the phenomenon which implies that the results are not driven by one particular set of occupations. Finally, the complementarity between advanced computer use and city size occurs within detailed occupations.[3] Whereas Berger and Frey (2016) suggest that cities with a large endowment of high-skilled workers benefit more from recent technological change, the current paper hints at a complementarity between computer and city size given the skill of workers. My main challenge though is to exclude unobserved skill as the driver of the complementarity.[4] The correlation is robust to the inclusion of additional workers characteristics such as the type of school attended or the highest math qualification. Also reassuring is the fact that contrary to sorting across city sizes according to observed skill, the urban economic literature has found that sorting according to *unobserved* ability is surprisingly low (see the introduction of De la Roca et al. (2018) for a review.)

Next, I show that the correlation between the probability to have an advanced use of computer at work is driven to zero when I control for direct measures of complex tasks at work from the German Working population Survey. I use an interactive tasks index and an analytical tasks index as well as the percentage time working on a computer. In the model the latter represents the share of time allocated to complex tasks. There is thus

[1]On system of cities, see the seminal papers of Henderson (1974), Henderson (1991) as well as Fujita (1989), Fujita and Thisse (2013), Glaeser (2008). See Behrens and Robert-Nicoud (2015) for a review and Allen and Arkolakis (2014) and Redding (2016) for recent contributions. Developing models with heterogeneous workers is important to be able to discuss issues of inequality and redistribution.

[2]Skill sorting across city size have been widely documented (E.g. Combes et al. (2008)) and tend to be increasing over time (Berry and Glaeser (2005)). For a review see the handbook chapter of Combes and Gobillon (2015).

[3]Arntz et al. (2016) documents the potential for tasks' automation within occupations. My analysis highlights the unevenness of this development across city size. Other papers discussing uneven organization of production across space include Rossi-Hansberg (2005), Tian (2018), Spanos (2019). See Redding and Rossi-Hansberg (2017) and Proost and Thisse (2019) for reviews.

[4]This concern relates to the one of DiNardo and Pischke (1997) when explaining higher wages for workers working with computers.

no within task variation in computer capital across city size. This might be an indication that tasks are at the core of the complementarity between computer capital and city size or that complex tasks captures unobservables related to the complementarity.[5]

Moreover, I dig into the following underlying mechanisms behind the empirical relationship between computer capital and city size. I first exclude that my complementarity is mainly explained by a labor capital complementarity in the presence of agglomeration economies. Indeed, I show that among the 14 equipment types I have available, only two display a positive relationship with city size given workers' characteristics. Differences in the size of equipment and infrastructures specific to large cities are related to the complementarity of computerized equipment and city size. Overall, wherever the origin of the complementarity, it is politically relevant. Indeed, if firms allocate more computerized equipment to a given worker in large cities, differences in urbanization rate can explain differences in investment in computer capital across country.[6]

Last, I regress wages on the triple interaction of city size, computer capital and skill. I find that an increase in city size of 10% is associated with an increase of 0.62% in the wage gap between workers in occupation requiring a university degree and occupations not requiring it, when both groups make an advanced use of computers at work. Although is remains a correlation, it suggests that computerization is a valid candidate to partly explain the choice of high skilled workers to live in large cities.

My paper also contributes to other strands of the literature. Existing economic literature has highlighted that IT technology is overall a substitute for face-to-face interactions and hence economic activities become more dispersed over space with recent technological change (Ioannides et al. (2008)). Consequently, concentration forces of technological change - for example an increase in overall contact regularity due to information technologies that can in turn also foster face-to-face contact (Gaspar and Glaeser (1998), Charlot and Duranton (2006), Leamer and Storper (2014) see Van Reenen et al. (2010) Section I.C for a review) - were offset by its dispersive forces. Those are acting through for example the possibility for workers to communicate over long distances. Nevertheless, I argue that the balance between concentration and dispersion forces linked to recent technological change depends on the skill level, where housing costs is a strong dispersive force for lower skilled workers.[7] Specifically, computerization is a complement to face-to-face interactions for workers with a strong comparative advantage with complex tasks. As it specifically raises their productivity in large cities recent technological change might consequently be "city-biased" for them on top of being generally "skill-biased".[8]

Finally, and in this context, I provide a new perspective on real wage inequalities.

[5]Direct measures of tasks positively relate to city size given workers' characteristics. Contrary to interactive tasks, analytical tasks positively relate to city size given detailed occupations. The current paper thus adds further evidence that tasks content of occupation as measured by the International Standard Classification of Occupations (ISCO) varies across city size (e.g. Kok (2014)).

[6]See Draca et al. (2006) for a review of the evidence about productivity differences linked to ICT between Europe and the US.

[7]See e.g. Gyourko et al. (2013) and Ganong and Shoag (2017) for a discussion of the difficulties of accessing large cities for them.

[8]For a recent discussion about the origin of the urban wage premium, see Combes et al. (2012), De La Roca and Puga (2017).

The literature on the impact of computerization on inequalities emphasizes the increasing demand for skilled labor following computerization that might lead to increasing wage inequalities if the supply of skilled labor does not rise significantly (e.g. Acemoglu and Autor (2011)). Nevertheless, most often overlooked in this debate and what my theoretical analysis suggests is that computerization has led to a reallocation of high-skilled workers in large metropolitan areas where living costs are highest and tend to increase. Consequently, nominal inequalities is larger compared to inequalities adjusted for purchasing power. This point has been highlighted by Moretti (2013) and recent empirical contributions about real wage inequalities and ability sorting in this context include Diamond (2016), Handbury (2013) and Wang (2016).[9] My approach suggests a different mechanism and thus complements theirs.

The rest of this paper is structured as follow. The next section formalizes the city-computer complementarity in a spatial equilibrium framework, Section 1.3 brings empirical support for the complementarity given detailed workers' and firms' characteristics. Next, Section 1.4 focuses on the potential mechanism behind the highlighted city-computer complementarity. Then, Section 1.5 exposes the correlation between individual wages and the triple interaction of city size, computer capital and education. Finally, Section 1.6 concludes.

1.2 Model

This section formalizes the complementarity between complex tasks, computerized equipment and city size in an urban model of a system of cities.

1.2.1 Endowment

The economy consists of a continuum of workers with heterogeneous skill levels s distributed between 0 and 1 and described by the density $g(s) \equiv L(s)/L$. $L(s)$ continuous in s denotes the population with skill s in the economy, with $\frac{\partial L(s)}{\partial s} < 0$. $L = \int_0^1 L(s)ds$ is total population there. Skill is defined in relation with the ability to perform complex tasks. The set $\mathcal{A} \subset \mathbb{R}^n$ is the set of productivities of the n discrete cities that can host the workers. I use the terms city and agglomeration interchangeably by meaning an integrated labor market. They exogenously differ in their productivity $A > 0$ which reflects for example natural advantages. The distribution of A is such that no two cities have the same A. Cities are endowed with a constant housing stock $\bar{H}$, owned by absentee landlords. All locations are homogeneous within a city (there are no commuting costs). The size and skill composition of cities are the main endogenous variables in the model.

Preferences

The preferences of the - perfectly mobile - workers depends on the consumption of a fully tradable final good C (chosen as the numéraire) and housing H. Denote by $w_A(s)$ the

[9] Diamond (2016) emphasizes that nominal inequalities *understate* real inequalities because of endogenous amenities who particularly benefit skilled people in large cities.

equilibrium wage earned in city A by a worker with skill s and by r_A the unit price of housing:

(i) Workers choose the amount of final good C and housing H which maximizes the following Cobb-Douglas utility function given the wage $w_A(s)$:

$$\max_{C,H} u(C,H) = \left(\frac{C}{\mu}\right)^{\mu}\left(\frac{H}{1-\mu}\right)^{1-\mu}$$

subject to the budget constraint $w_A(s) = C + r_A H$ and $\mu \in (0,1)$.

(ii) Workers choose a location A to maximize indirect utility $V_A(s)$:

$$\max_{A} V_A(s) = r_A^{\mu-1} w_A(s) \tag{1.1}$$

The following market clearing condition for housing holds in each city:

$$\int_0^1 H_A(s)L_A(s)ds = \overline{H},$$

where $H_A(s) = \frac{1-\mu}{r_A} w_A(s)$ is the equilibrium housing demand as a function of the skill level. The expression for the equilibrium rent in city A is given by:

$$r_A = (1-\mu)\frac{L_A}{\overline{H}}\bar{w}_A \tag{1.2}$$

The rent positively depends on city size L_A and on the average wage in the city $\bar{w}_A \equiv \int_0^1 w_A(s)g_A(s)ds$.

The determination of the wage in a city $w_A(s)$ is discussed in the next subsections.

Technology

Production of the numéraire good, denoted as Y_A, in a city takes place in a perfectly competitive environment.[10] The final good is produced by a representative firm. The firm has four choice variables for each skill unit s: inputs in computer capital $K_{Z,A}(s)$, in complex tasks $Z_A(s)$ and simple tasks $X_A(s)$, as well as a second type of capital, used in complementarity with simple tasks $K_{X,A}(s)$.[11]

$$Y_A = A\left\{\int_0^1 \phi_Z(s,L_A)Y_{Z,A}(s)^{\frac{\beta-1}{\beta}} + \phi_{X,A}(s)Y_{X,A}(s)^{\frac{\beta-1}{\beta}}ds\right\}^{\beta/(\beta-1)} \tag{1.3}$$

[10] This assumption together with firm homogeneity and constant returns to scale allow me to focus on a representative firm in each city. Indeed, as profit is zero everywhere firms are indifferent across locations. The problem of sorting of firms across locations is thus greatly simplified.

[11] The treatment of complex versus simple tasks is based on Peri and Sparber (2009).

where $Y_{Z,A}(s) = K_{Z,A}(s)^{\alpha} Z_A(s)^{1-\alpha}$ is the complex output delivered by the skill unit and $Y_{X,A}(s) = K_{X,A}(s)^{\alpha} X_A(s)^{1-\alpha}$ its simple output. Computer capital is thus complement with complex tasks and simple capital is complement with simple tasks. Moreover, $\beta > 1$ is the constant elasticity of substitution between skill units and types of inputs.

Assumption 1 relates to the productivity of the simple tasks, simple capital unit $\phi_{X,A}(s) > 0$, and the productivity of the complex tasks, computer capital unit $\phi_Z(s, L_A) > 0$.

Assumption 1 *(i)* $\phi_{X,A}(s) = 1$, *constant* $\forall s, A$;

(ii) $\frac{\partial \ln \phi_Z(s,L_A)}{\partial s} > 0, \ \frac{\partial \ln \phi_Z(s,L_A)}{\partial L_A} > 0$;

(iii) $\frac{\partial^2 \phi_Z(s,L_A)}{\partial^2 s} < 0, \ \frac{\partial^2 \phi_Z(s,L_A)}{\partial^2 L_A} < 0$ *with* $\frac{\partial \ln \phi_Z(s,L_A)}{\partial L_A} < \frac{1}{L_A}$;

(iv) $\lim\limits_{L_A \to 0} \dfrac{\partial \phi}{\partial L_A} \to \infty, \ \lim\limits_{L_A \to L} \dfrac{\partial \phi}{\partial L_A} \to 0$;

(v) $\frac{\partial^2 \ln \phi_Z(s,L_A)}{\partial L_A \partial s} \geq 0$.

To begin with, parts (i) and (ii) of Assumption 1 reflect the key sources of comparative advantage of the model. First, higher skilled workers have a comparative advantage performing complex tasks relative to simple ones where the productivity $\phi_Z(s, L_A)$ of the capital-complex tasks unit $Y_{Z,A}(s)$ is increasing in the skill s of the worker.[12] For simplicity, I assume that the productivity of simple tasks is independent from the skill s of workers. Second, large cities have a comparative advantage performing complex tasks relative to simple ones where the productivity of complex tasks $\phi_Z(s, L_A)$ positively depends on the population in the city. It captures agglomeration or external scale economies and is treated as given by the firm. I also assume for simplicity that the productivity of simple tasks is independent from city size. Nevertheless, the productivity of simple tasks could also be positively related to city size, where city size would have an absolute advantage in performing both tasks. Key is the comparative advantage with complex tasks. The rationale behind this assumption is that complex tasks benefit from city size through knowledge spillovers on top of other agglomeration mechanisms. Simple tasks do not.[13] Indeed, as part of the knowledge to perform complex tasks is tacit, one cannot write a computer code to do it. Similarly, tacit knowledge implies that one cannot learn it on her own with books as it is not codified. As a result, one need to learn from other skilled people and thus especially benefit from the environment of large cities.[14]

[12] See the review paper of Acemoglu and Autor (2011) .

[13] Previous literature documented a higher return to analytical skill in larger cities relative to smaller ones (Bacolod et al. (2009) and Florida et al. (2011)). See Duranton and Puga (2004) for an overview of agglomeration mechanisms.

[14] For a discussion of tacit knowledge and its transmission within the firm see Polanyi (2009), first published in 1966, Foray (2004), Thoenig and Verdier (2010) among others.

Next, part (iii) and (iv) of Assumptions 1 encompasses regularity assumptions, which ensures that the problem is concave, as well as Inada conditions. Finally, part (v) of Assumptions 1 states that the cross-derivative of the logarithm of the productivity of the capital-complex tasks unit $Y_{Z,A}(s)$ must be non-decreasing in s and L_A. Note that the inequality need not be strict.

Moreover, the skill of workers encompasses both vertical and horizontal differentiation. Indeed, higher skill levels are more productive at performing complex tasks and also different skill levels have specific skills. Because labor work is non-tradable across cities, this implies that all skill levels are needed to produce in a city (the support of s is thus the same in all cities). However, there is an amount of substitutability between skill units captured by the parameter $\beta > 1$. Consequently, the labor market in which workers compete in the city is "skill-specific". Whereas the local labor market in which the skill types compete is skill-specific, agglomeration economies impact all skill types.

The firm minimizes costs $Costs_A = p_Z \int K_{Z,A}(s)ds + p_Z \int K_{Z,A}(s)ds + \int w_{z,A}(s)Z_A(s)ds + \int w_{x,A}(s)X_A(s)ds$ for a given level of output Y_A, where $w_{x,A}(s)$ is the wage rate for complex tasks, $w_{x,A}(s)$ the wage rate for simple tasks, $p_Z > 0$ the price of computer capital $K_{Z,A}(s)$ and $p_X > 0$ the price of simple capital $K_{X,A}(s)$ which are both supplied perfectly elastically at a constant prices p_X and p_Z. The Lagrange multiplier of the firm minimization problem, which is equal to the marginal cost, is equal to the price of the final good in equilibrium.

The last element of the model to be discussed is the worker's allocation choice of her working time across tasks. Each worker inelastically supplies one unit of labor which can be divided between complex $l_{z,A}(s)$ and simple tasks $1 - l_{z,A}(s)$ where $l_{z,A}(s)$ is the share of working time spend on complex tasks. Even though labor supply is inelastic given the choice of the city, perfect labor mobility across cities leads to a perfectly elastic labor supply in spatial equilibrium. The worker chooses $l_{z,A}(s)$ in order to maximize her wage $w_A(s)$ given the choice of the city, where $w_{z,A}(s)$ is the wage rate for complex tasks and $w_{x,A}(s)$ the wage rate for simple tasks:

$$\max_{l_{z,A}(s)} w_A(s) = l_{z,A}(s)^{\delta} w_{z,A}(s) + [1 - l_{z,A}(s)]^{\delta} w_{x,A}(s) \tag{1.4}$$

Worker's output of complex and simple tasks are $z_A(s) = l_{z,A}(s)^{\delta}$ and $x_A(s) = [1 - l_{z,A}(s)]^{\delta}$, respectively. Lower case letters denote per worker variables, $z_A(s)$ being for example the amount of complex tasks supplied by a worker with skill s in city A. $\delta \in (0, 1)$ implies decreasing returns to the type of tasks.[15] As a result, an internal solution exists, every type of workers spend time on both types of tasks and the problem is concave, so the second order conditions are satisfied.

[15] Peri and Sparber (2009) illustrate this assumption by taking the example of a researcher writing a complex paper for most of his day where it is efficient for him to allocate a little of his time on simple tasks like cleaning up his desk.

1.2.2 Equilibrium share of time allocated to complex tasks and ratio of computer to simple capital assigned to workers given city population

I first consider the equilibrium given the choice of the city and then describe the spatial equilibrium. City size L_A and the skill composition in the city $g_A(s)$ are thus given in this subsection as well as in Subsection 1.2.3. I leave some of the calculations for Appendix A.1.

The equilibrium amount of time allocated to complex tasks is given by:

$$l_{z,A}(s) = \frac{B_{Z,A}(s)}{1 + B_{Z,A}(s)}. \tag{1.5}$$

where $B_{Z,A}(s) = \phi_Z(s, L_A)^{\frac{\beta}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}} \left[\frac{p_X}{p_Z}\right]^{\frac{\alpha(\beta-1)}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}}$.

Because of decreasing returns to tasks ($\delta < 1$) and because $\phi_Z(s, L_A)$ is strictly positive, every worker spend a positive amount of time on each type of tasks. The restrictions $\frac{\partial^2 \phi_Z(s,L_A)}{\partial^2 L_A} < 0$ and $\frac{\partial^2 \phi_Z(s,L_A)}{\partial^2 s} < 0$ ensure that $B_{Z,A}(s)$ remains bounded for all possible s and L_A. Note that the optimal amount of time allocated to complex tasks is independent from the local labor supply $L_A(s)$, and that $a \equiv \frac{\beta}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta} > 1$. It is increasing in $\phi_Z(s, L_A)$ and in $\frac{p_X}{p_Z}$.

Lemma 1 *The optimal share of time allocated to complex and simple tasks behaves as follows:*

(i) If $\frac{\partial^2 \ln \phi_Z(s,L_A)}{\partial L_A \partial s} = 0$, *then* $\frac{\partial^2 \ln l_{z,A}(s)}{\partial L_A \partial s} < 0$; *if* $\frac{\partial^2 \ln \phi_Z(s,L_A)}{\partial L_A \partial s} > 0$, *then* $\frac{\partial^2 \ln l_{z,A}(s)}{\partial L_A \partial s}$ *is ambiguous.*

(ii) $\frac{\partial^2 \ln\left(1-l_{z,A}(s)\right)}{\partial L_A \partial s} < 0$.

Proof. See Appendix A.1.2.

The optimal share of time allocated to complex tasks as a function of city size and skill depends on two opposing effects. First of all, the percentage increase in the share of time allocated to complex tasks with city size tends to be decreasing in the skill of workers. Indeed, time available to workers is limited and returns to tasks are decreasing ($\delta < 1$). Because high-skilled workers spend a larger share of time on complex tasks regardless of location, their possibility to increase it with city size is more limited compared with low-skilled workers. If $\frac{\partial^2 \ln \phi_Z(s,L_A)}{\partial L_A \partial s} = 0$, this is the only effect differentiating skill types. Second, if the productivity of the computer capital - complex tasks unit $\phi_Z(s, L_A)$

is log supermodular in s and L_A, the percentage increase in the share of time allocated to complex tasks with city size tends to be increasing in the skill of workers. The net effect in this latter case is theoretically ambiguous. Part (i) of Lemma 1 summarizes this finding.

Part (ii) of Lemma 1 states that the percentage decrease in the share of time allocated to simple tasks with city size is larger for high-skilled workers. Because high-skilled workers have an overall lower share of time allocated to simple tasks, the denominator is lower to compute a *percentage* decrease.[16]

Next, the equilibrium ratio between complex and simple capital per worker given by:

$$\frac{k_{Z,A}(s)}{k_{X,A}(s)} = \left(\frac{l_{z,A}(s)}{1-l_{z,A}(s)}\right)^{\frac{\delta(1-\alpha)(\beta-1)}{\beta(1-\alpha)+\alpha}} \left(\frac{\phi_Z(s,L_A)P_X}{p_Z}\right)^{\frac{\beta}{\beta(1-\alpha)+\alpha}}$$

$$\leftrightarrow \frac{k_{Z,A}(s)}{k_{X,A}(s)} = \phi_Z(s,L_A)^{\frac{\beta}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}} \left[\frac{p_X}{p_Z}\right]^{\frac{\beta(1-\delta+\alpha\delta)+\delta(1-\alpha)}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}} \tag{1.6}$$

The ratio between complex and simple capital is increasing in its relative productivity $\phi_Z(s,L_A)$ and decreasing in its relative price $\frac{P_Z}{P_X}$. Moreover, it is independent from the local labor supply of the skill $L_A(s)$.

1.2.3 Indirect utility and city population

Wage given city population

The equilibrium wage given city population is specified as follow (see Appendix A.1 for derivation, using equation (1.4)):

$$\begin{aligned} w_A(s) =& A^{\frac{1}{1-\alpha}+\frac{\beta-1}{\beta(1-\alpha)+\alpha}} \left[\frac{\alpha}{p_X}\right]^{\frac{\alpha}{1-\alpha}+\frac{\alpha(\beta-1)}{\beta(1-\alpha)+\alpha}} L_A(s)^{\frac{-1}{\beta(1-\alpha)+\alpha}} (1-\alpha)\left[1+B_{Z,A}(s)\right]^{\frac{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}{\beta(1-\alpha)+\alpha}} \\ & \left\{\int_0^1 L_A(s)^{\frac{(\beta-1)(1-\alpha)}{\beta(1-\alpha)+\alpha}} \left[1+B_{Z,A}(s)\right]^{\frac{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}{\beta(1-\alpha)+\alpha}} ds\right\}^{\frac{\beta(1-\alpha)+\alpha}{(1-\alpha)(\beta-1)}} \end{aligned} \tag{1.7}$$

As $\frac{1}{1-\alpha}+\frac{\beta-1}{\beta(1-\alpha)+\alpha}>0$, the wage of a worker with skill s - all else equal, is increasing in the exogenous productivity of the site A by inspection.

The following lemma summarizes two key properties:

[16]Focusing on the share of time allocated to complex and simple tasks instead of their log shows perfect symmetry between the two.

Lemma 2 *The wage of a worker with skill s - all else equal:*

(i) is log supermodular in s and L_A (local labor supply $L_A(s)$ is given): $\frac{\partial^2 \ln w_A(s)}{\partial s \partial L_A} > 0$, $\forall s$.

(ii) The log supermodularity of the wage of the worker in s and L_A increases as the relative prices of computerized equipment p_Z relative to simple equipment p_X decreases, for all parameter values.

Proof. See Appendix A.1.3.

Part (i) of Lemma 2 reflects part (ii) of assumption 1. Due to the fact that higher skilled workers have a comparative advantage performing complex tasks, they allocate a larger share of time to them, which implies that a larger proportion of their productive time benefits from agglomeration economies. Note that this is true for all s, and also true when $\frac{\partial^2 \ln \phi_Z(s,L_A)}{\partial L_A \partial s} = 0$.

The key property for my argument that computerization is a micro foundation of the skill-bias of agglomeration economies is then part (ii) of Lemma 2. As the price of computerized equipment decreases relative to simple equipment, the log supermodularity of the wage of the worker in s and L_A increases.[17] As part (iii) of Lemma 2, this property is also true when $\frac{\partial^2 \ln \phi_Z(s,L_A)}{\partial L_A \partial s} = 0$. As the relative price of computerized equipment decreases, its investment increases, and especially so in large cities where its productivity is comparatively large. Because high-skilled workers allocate more time to complex tasks, complementary with computerized equipment, a larger proportion of their working time benefits from the additional productivity in large cities due to the additional investment in computerized equipment there. Consequently, the skill-bias of agglomeration economies increases.

Rent and city population

The land rent in the city (equation (1.2)) is increasing in the average wage there as well as city population:

$$r_A = \frac{(1-\mu)}{\overline{H}} A^{\frac{1}{1-\alpha}+\frac{\beta-1}{\beta(1-\alpha)+\alpha}} \left[\frac{\alpha}{p_X}\right]^{\frac{\alpha}{1-\alpha}+\frac{\alpha(\beta-1)}{\beta(1-\alpha)+\alpha}} (1-\alpha) \left\{\int_0^1 L_A(s)^{\frac{(\beta-1)(1-\alpha)}{\beta(1-\alpha)+\alpha}} \left[1+B_{Z,A}(s)\right]^{\frac{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}{\beta(1-\alpha)+\alpha}} ds\right\}^{\frac{(1-\alpha)(2\beta-1)+\alpha}{(1-\alpha)(\beta-1)}} \tag{1.8}$$

[17]The decrease in the price of computerized equipment relative to other equipment reflects the continuous development of the efficiency of computers. With the development of the semiconductor technology, the material needed to perform a given tasks on a computer decreases. This was theorized as Moore's law which states that the number of components per chip is approximately doubling every year (see Mack (2011) for an overview of 50 years of Moore's law).

Indirect utility and city population

The log supermodularity in city size and the skill of the wage (Lemma 2) translates to the indirect utility because housing costs do not depend on the skill level and preferences are homothetic. The no-black hole condition $\mu < \left[2 + \frac{1}{(1-\alpha)(\beta-1)}\right]^{-1} \in \left(0, \frac{1}{2}\right)$ is necessary to prevent a solution where the whole population agglomerates in one single city. That is, the preferences of workers for the final good must not be too large.

Lemma 3 *For a given skill distribution, the utility of a worker is log supermodular in s and L:* $\frac{\partial^2 V_A(s)}{\partial s \partial L_A} > 0$ *and the log supermodularity increases as the relative prices of computerized equipment* p_Z *relative to simple equipment* p_X *decreases:* $\frac{\partial^3 V_A(s)}{\partial s \partial L_A \partial \frac{p_Z}{p_X}} < 0$.

Proof. In the text.

1.2.4 Spatial equilibrium

This section deals with the endogenously determined population and skill composition in cities. It further asks what are its implications for the assignment of computer capital to workers as the exogenous productivity of the site varies. Variables with a subscript E denote spatial equilibrium variables. I focus on an equilibrium where all cities are populated.

Definition of spatial equilibrium

A spatial equilibrium is an allocation of population to cities $L_{E,A}(s)$ such that for all cities:

(i) Utility is equalized across cities:

$$V_A(s) = V_E(s) \ \forall \ A, s \tag{1.9}$$

same $V_E(s)$ where $V_A(s)$ is given by equation (1.1);

(ii) The whole population is allocated to the n cities:

$$\sum_{A \in \mathcal{A}} L_{E,A}(s) = L(s) \tag{1.10}$$

(iii) Workers maximize utility and firms maximize profits.

Local wages $w_{E,A}(s)$, rents $r_{E,A}$ and the constant level of utility across cities $V_E(s)$ support the equilibrium allocation.

Moreover, a spatial equilibrium needs to be stable to small perturbations. Since the equilibrium concept is a decentralized one, I define stability with respect to perturbation of the location choice of a single type of workers (and not with respect to a perturbation of the whole assignment function $L_{E,A}(s))$[18]:

$$\frac{\partial V_A(s)}{\partial L_{E,A}(\check{s})} \leq 0 \; \forall \check{s}; \forall s.$$

The utility of any workers with skill s must be decreasing or be unaffected if a new inhabitant comes to the city in equilibrium whatever the skill $\check{s}$ of the new comer.

Properties of spatial equilibrium

Proposition 1 *:*

(i) Population increases with the exogenous component of the productivity of the site A: $\frac{\partial L_E(A)}{\partial A} > 0$*;*

(ii) The skill distribution of high-A cities first order stochastically dominates the one of low-A cities: $\frac{\partial^2 \frac{L_E(s,A)}{L_E(A)}}{\partial A \partial s} > 0$*;*

(iii) The representative firm allocates more computerized equipment to a worker of a given skill in large cities compared to smaller ones: $\frac{\partial k_{Z,E}(s,A)}{\partial A} > 0$*;*

(iv) Simple capital allocation with city size is theoretically ambiguous: $\frac{\partial k_{X,E}(s,A)}{\partial A} \gtrless 0$*, But:* $\frac{\partial k_{Z,E}(s,A)}{\partial A} > \frac{\partial k_{X,E}(s,A)}{\partial A}$.

Proof. See Appendix A.1.6 and in the text.

Part (i) of Proposition 1 states that more productive sites are more populated. On the one side, indirect utility is increasing in A all else equals for all skill types as is their wage. On the other side, the stability condition and the second order conditions imply that an increase in population decreases indirect utility. Consequently, an increase in population with the exogenous productivity A of the site is a necessary condition to balance the effect of a higher exogenous productivity on utility **for all skill types** and have a constant utility across sites in equilibrium.

Furthermore, part (ii) of Proposition 1 highlights that more productive sites host a more skilled population. This bears on the result that high-skilled workers disproportionately benefit from larger cities given the skill composition (Lemma 3) and that the local labor demand is downward sloping. Real wages can thus be equalized for all s and all A.

Next, part (iii) of Proposition 1 relates to the increasing assignment of computerized equipment as a function of the exogenous productivity of the site A. The result reflects

[18]Stability with respect to the whole assignment function would though allow to reduce the number of equilibria as all equilibria with unpopulated site would be unstable.

the complementary between city size and computerized equipment. The latter stems directly from $\phi_Z(s, L_A)$ depending on L_A and indirectly through local GDP Y_A.

Moreover, part (iv) of Proposition 1 shows that the behavior of simple capital with city size is ambiguous. Indeed, it tends to be increasing in city size through higher local amenity A and more productive workers there (higher local output even for a given A). Conversely, its complementarity with simple tasks goes in the opposite direction. Equation (1.6) implies that the relationship of computer capital with city size is in any case stronger than the one of simple capital and city size: $\frac{\partial k_{Z,E}(s,A)}{\partial A} > \frac{\partial k_{X,E}(s,A)}{\partial A}$.

Finally, note that the behavior with skill of the elasticity of computer capital's allocation with city size is not a distinguishing feature of my model. Indeed, it might either be increasing or decreasing in skill if $\frac{\partial^2 \ln \phi_Z(s,L_A)}{\partial L_A \partial s} > 0$: $\frac{\partial^2 \log k_Z}{\partial s \partial A} \gtrless 0$ (**Proof.** See Appendix A.1.6). It mirrors Lemma 1 in Section 1.2.2 as the elasticity of computer capital with city size is equal to the elasticity in time allocated to complex tasks with city size. The link between the 2 first comes from the optimality condition for computer capital (equation (A.17)), where the optimal amount of computer capital positively depends on the share of time allocated to complex tasks to the power of delta. The second link comes from the spatial equilibrium condition. The allocation of computer capital to workers in a skill group also *negatively* depends on the local share of workers in that group. This reduces the considered elasticity up to the point where it follows the pattern of the share of time allocated to complex tasks.

Proposition 2 *The proportion of high-skilled workers in large cities increases as the ratio of prices of simple to complex capital p_X/p_Z increases:* $\frac{\partial^3 \frac{L_E(s,A)}{L_E(A)}}{\partial A \partial s \partial p_X/p_Z} > 0$.

Proposition 2 is the key statement of this paper. It follows from part (ii) of Lemma 2: the log supermodularity of the wage - *given city population* - increases when the relative price of computer capital relative to simple capital decreases. Nevertheless, in spatial equilibrium, because of perfect mobility and skill complementarity, the wage premium of high skilled workers in large cities is driven to zero. As profits for firms' entry in a perfect competition framework, wages induce the sorting of high-skilled workers in large cities and are absent at the equilibrium.

Proposition 1 and 2 imply the following for local wages $w_{E,A}(s)$ and rents $r_{E,A}$ which support the equilibrium allocation. First of all, the rent $r_{E,A}$ is increasing in the exogenous productivity of the city A since it positively depends on the average wage $\bar{w}_{E,A}$ and the population $L_{E,A}$ in the city (part (i) and (ii) of Proposition 1). Furthermore, the latter assertion in turn implies that the wage $w_{E,A}(s)$ increases with the exogenous productivity of the site for all skill levels. Finally, a total population $L(s)$ decreasing in s implies that $V_E(s)$ is monotonically increasing in s. This latter condition also ensures that wages as well as computer capital's allocation are increasing in s in a given location.

Appendix A.1.5 shows the existence of an equilibrium, using equation (1.9) and 1.10 to obtain an explicit system of differential equations. Moreover, it discusses the conditions for equilibrium stability.

The rest of this paper is organized as follow. First, Section 1.3 test part (iii) of Proposition 1. Indeed, if it fails to be verified for all skill groups, my model is not valid. Appendix A.4 shows the behavior of this elasticity with skill types, even though it is not a distinguishing feature of my model.

Second, Section 1.4 details possible mechanisms underlying the complementarity between city size and computer capital. I first focus on complex tasks as being at the heart of the complementarity as highlighted by this section. Second, I inquire whether labor capital complementarities in the presence of agglomeration economies is the main explanation behind the complementarity. This mechanism is also included in my model but affects all types of capital and lies behind the ambiguity of the relationship of simple capital and city size (Part (iv) of Proposition 1). If capital that can be characterized as simple does not display a positive relationship with city size given the skill of workers, I can exclude this mechanism as the main driver of the computer capital-city size complementarity. Finally, I present competing mechanisms neglected by the present model such as difference in the size of equipment.

Last, Section 1.5 focuses on the wage behavior associated with computer capital and city size. The technology in equation (1.3) implies that allocating computer capital disproportionately increases the productivity of high-skilled workers in large cities.[19] In the model, though, both computer capital and city size are endogenously determined. While I have an instrument available for city size (long-lagged population), I do not have an instrument for computer capital. I thus regress - as a correlation - wages on the triple interaction of city size, computer capital and skill. Because there is no wage premium for high-skilled in large cities at equilibrium in the model, this exercise can be interpreted as relaxing my mobility assumption. First of all, computerization is an ongoing phenomenon and adjusting to it might take time. If wages adjust quicker than labor is mobile, a positive relationship of wages, cities and skill emerges as a consequence. Second, other spatial sorting mechanisms neglected in the present model might also be at play.

1.3 Testing the complementarity between computer capital and city size

This section provides some evidence for prediction (iii) of Proposition 1 which states that firms allocate more computerized equipment to workers of a given skill type as city size increases.

[19]Equation (1.3) is at the core of Lemma 2 and Proposition 2.

1.3.1 Empirical strategy

In an ideal world, in order to test part (iii) of Proposition 1, I would run the following regression, where I expect $\theta_1 > 0$:

$$ComputerCapital_i = \theta_0 + \theta_1 A_j + \theta_2 Skill_i + \epsilon_i \tag{1.11}$$

where computer capital is measured for individual i, A_j is the natural advantage of the agglomeration j in which she lives and her skill $Skill_i$ is a comprehensive measure of her ability to perform complex tasks. Agglomerations are based on commuting flows. I depart from this ideal in three respects.

First, I do not have a perfect measure of the computer capital assigned to a worker in a given location. Investment in computer capital is usually available at the firm level, not at the worker level. Moreover, high-depreciation rates and price deflators makes it particularly difficult to measure capital in the ICT sector (see Draca et al. (2006) for a review of measurement issues specific to computer capital). I instead proxy computer capital with its use by the worker, which includes for example word-processing or analyzing data. I assume that a more advanced use of computer comprehends a larger investment in computer capital as it requires more computer power. Because I construct a dummy variable as my benchmark dependent variable, I will run a non-linear version of equation (1.11).

The second departure from equation (1.11) is that I do not observe the natural advantage A of an agglomeration. Based on prediction (i) of Proposition 1, which states that cities with larger A receive more inhabitants, I use city size Pop_j instead. In this context, I cannot interpret its coefficient as the causal effect of city size on the allocation of computer capital, because city size L_A is endogenously determined. As a result computerization could cause city size to increase rather than the other way around. For example, if computer capital disproportionately increases labor productivity in flat location because, say, of better internet access for a given investment, computerization can be one cause of the larger population in those location, because of increasing wage there. I thus also instrument city size in equation (1.11) with longed-lagged population.[20,21] The main focus on this chapter remains though the correlation between computer capital and city size.

The third departure from equation (1.11) is that I do not have a comprehensive measure for the skill of workers. Moreover, as I do not observe the same workers at different points in time in different cities, I cannot control for workers' fixed effects. Instead, I compare workers with the same observed individual characteristics Ind_i, including having a university degree, the type of school attended or highest math qualification. My

[20] Standard instrument in the urban economic literature, see Combes and Gobillon (2015), Section 4.3.2. Appendix A.5 describes the construction of the instrument.

[21] Knowing the causal impact of population size on computer capital is of interest for example to discuss the possibility that difference in urbanization rate across countries can lead to differences in per worker investment in computer capital, and thus to differences in per worker productivity, given their skill (See Draca et al. (2006) for a review of the evidence about productivity differences linked to ICT between Europe and the US).

specification thus correctly identifies the correlation between computer capital and city size for a given skill provided that the sorting of workers according to their skill across location only occurs according to the observed characteristics available in my datasets. Reassuring in this respect is that the urban economic literature has found that sorting according to *unobserved ability* is surprisingly low (see the introduction of De la Roca et al. (2018) for a review.) Additionally, I use occupations and industries dummies as alternative measures of skill. Indeed, they partly capture the skill of workers, although part of the effect I am interested in can act through them as they are endogenous.

The final specification I use in this section takes the following form:

$$P(c_i = 1|x_i) = G(\gamma_0 + \gamma_1 Pop_j + \gamma_2 Ind_i) \tag{1.12}$$

where c_i is a binary variable which equals 1 if worker i has an advanced used of computer at work, which means that the latent variable related to her productivity when having an advanced used of computer at work c_i^* is above firms' cost τ for equipping the worker:

$$c_i = \begin{cases} 1 & \text{if} \quad c_i^* \geq \tau \\ 0 & \text{if} \quad c_i^* < \tau \end{cases}$$

and $c_i^* = \gamma_0 + \gamma_1 Pop_j + \gamma_2 Ind_i - \varepsilon_i$; $G(.)$ is the logistic cumulative distribution function and the cdf for ε_i; Pop_j is either a dummy variable "Large cities" for Germany or a continuous measure of size for the Great Britain, controls Ind_i include observed worker's characteristics, such as education and occupation.

1.3.2 Data and measurement

Benefits and drawbacks of the data sets

First of all, I use the 2006th wave of the German working population survey conducted by the Federal institute for vocational education and training (BIBB) and the Federal Institute for Occupational Safety and Health (BAuA) (Hall and Beerman (2011)). The survey offers detailed information about the job content of the working population in Germany and with $20'000$ observation is representative at the level of the agglomeration (97 planning region in Germany in 2006).[22] It offers information on the use of computerized equipment as well as other equipment type for comparison. The measure of skill includes 4 dummies for education as well as overall score on school leaving certificates for about 40 percent of the sample. Also, I only have access to the agglomeration variable in 20

[22]The smallest planning region (Raumordnungsregion) receives about 60 observations randomly drawn. See Hartmann (2009) and Hartmann (2006) for more details about the sampling method and description of the survey.

categories. The survey has been widely used at the national level to measure the tasks content of occupations (see e.g.Spitz-Oener (2006)). Kok (2014) uses the survey at the city level and shows the representativeness of the survey in terms of observations across city sizes' categories. Rather than agglomeration categories, she uses a dummy variable indicating whether a worker lives in a city (administrative definition) with at least $50,000$ inhabitants.

Second, I use the 2006th wave of the British Skill Survey (BSS) which includes detailed information about the job content of a representative sample of $7,787$ workers aged $20-65$ in the United Kingdom (Felstead et al. (2014)). The author of the survey have demonstrated its consistency with national aggregates from larger representative survey in terms of age groups, sex, education, occupation and industry (Felstead et al. (2007)).[23] I know the agglomeration of the respondent (Travel to work areas, 308 areas in the UK in 2006) and can thus construct a continuous measure of city size. Moreover, I have information on hourly wages of workers as well as questions on the type of school attended and highest math qualification, additionally to 5 education dummies. Nevertheless, the survey only possesses information on computerized equipment and some small agglomerations have no or too few observations to be representative of the average worker there. Table A.1 in Appendix A.2 compares the number of observations for 7 population categories of agglomerations (Travel to work areas). Moreover, I check the consistency of the survey by comparing descriptive statistics by categories of city sizes with the larger Annual Survey of Hours and Earnings (See Appendix A.3).[24]

Overall, the evidence presented in this section is robust to both datasets and their limitations. Appendix A.2 discusses the construction and representativeness of the samples.

Measure of computer capital

I use as a proxy for computer capital the type of work performed on computer. It is a measure of the intensive margin of computer capital as a more advanced use of computer requires the investment in more expensive softwares and hardwares. For Germany, I define a dummy variable "Advanced use of computer" which takes value 1 if the worker writes programs or macros as well as if she programs for others. I define the same variable for Great Britain, where an advanced use includes "using a computer for analyzing information" or "using computer syntax and/or formulae for programming", whereas a non-advanced use includes "printing out an invoice in a shop", "using a computer for word-processing" or "communicating by e-mail".

Since the question asked to the survey respondents exposes detailed computer usage measurement error should remain a minor issue. A further advantage of this measure is that it offers a wide source of variation that can be used for identification. Alternative measures of computer capital as having a computer at work lacks variation as the vast majority of workers do use a computer. Section A.8.1 focuses on "Having a computer at work" as a proxy for computer capital for comparison. Moreover, Appendix A.6 treats

[23]Men and young workers are slightly underrepresented, which they correct using weigths.

[24]To my knowledge there is no published work based on an analysis with TTWA in the BSS.

the variable on the use of computer as an ordered variable with four possible values: "No computer"; "Basic use of computer"; "Writing programs for oneself"; "Writing programs for others". It shows that workers without a university degree have a larger probability to have a basic computer use at work when city size increases compared to not using a computer. Having a computer at work is thus a useful measure for the variation in computer capital for them. On the contrary, the probability that workers with a university degree have a basic computer use decreases with city size to the benefit of more advanced use. Indeed, almost all of them work with a computer. Having a computer at work or not is thus not a useful measure for the variation in computer capital for them.

Measure of city size

I use the German "planning regions" as my functional definition for labor markets.[25] The definition available in 2006 is constructed from the German NUTS-3-regions, which represents about 400 districts with a minimum of 150′000 inhabitants: one region is added to the core one if commuting outflows from that region to the core region is above 15% of the working force and corresponds to the boundaries of states with one exception. The 97 planning regions cover the whole German territory. I have their density in 20 categories (population/km^2 according to 2006 figures).[26] As agglomerations are constructed from relatively large spatial units, I prefer to measure their size with population density rather than overall population.

For the analysis below I define a dummy variable "Large city" if the agglomeration pertains to large urban centers as defined by the German administration and zero otherwise. The German administration defines urban centers mainly with regard to population density (about more than 274 inhabitants per km^2) as well as the presence of a large city center.[27]

For the case of Great Britain, I measure city size as a continuous variable with the population of the travel to work area (TTWA) according to population census of 1991.[28] TTWA is a functional definition of an integrated labor market where 75% of the workforce of an area must live in the area and 75% of the residents must work there. The Office for National Statistics uses about 10′000 wards to construct the 308 TTWAs, which cover the whole UK territory. It requires a viable TTWA as having at least 3500 inhabitants.[29]

[25] There are 2 functional definitions of labor market in Germany, the planning regions I use here "Raumordnungregionnen", and "Arbeitsmarktregionnen", constructed from smaller units.

[26] Appendix A.7 describes the density range of the categories of city density.

[27] http://www.bbsr.bund.de/BBSR/DE/Raumbeobachtung/Raumabgrenzungen/StadtLandRegionen _Typen/StadtLandRegionen _Typen.html?nn=443270, visited the 14th of January 2017.

[28] Source: Office for National Statistics and National Records of Scotland. The 1998 definition of TTWAs that was available in 2006 is based on commuting flows from 1991.

[29] Definition of TTWA in 1991 and 2001: see Bond and Coombes (2007), Office for National Statistics. The TTWA of the respondents is included in the British Skill Survey.

Measures of workers' characteristics

Finally, available observable workers' characteristics for Germany comprises: 4 education dummies including primary education, secondary 1 and 2, secondary 3 without a university degree and secondary 3 with a university degree; occupation as measured by the International Standard Classification of Occupations (ISCO) from 1988 at the 4 digit level; the 2 digit level industries (German Classification of Economic Activities, Edition 2003 (WZ 2003)); age and sex. I also use the first digit of occupation[30] as an alternative measure for education based on the International Standard Classification of Education (ISCED) required for the job.[31] Last, I have overall score on school leaving certificates for about 40 percent of individuals in the sample (very good, good, satisfactory, fair/pass). This score offers a wide variation within education groups.

For Great Britain, beside sex and age, I use five education dummies according to the National Vocational Qualification level[32], as well as the type of school last attended, including private, grammar public, comprehensive public, secondary modern school and city technology college (whole sample). Moreover, I also use highest math qualification (3 dummies related to grades for 65% of the sample). Occupation is defined at the 4 digit level according to the International Standard Classification of Occupations (ISCO) from 1988 and industry is available at the 4 digit level according to the UK Standard Industrial Classification of Economic Activities (SIC) definition from 1992.

Descriptive statistics

Table 1.1 compares the probability to have an advanced computer use across German and Great Britain datasets. As the education system slightly differs between theses two countries, I focus on the education required by the job rather than formal education (the prevalence of university degrees is stronger in Great Britain). "Large cities" are urban centers as defined by the German administration[33] and agglomerations with more than 440′000 inhabitants in Great Britain. The frequency of advanced computer use is slightly higher in Germany relative to Great Britain. In both cases, its mean is significantly larger in large urban centers relative to the rest of the country. Table A.14 in Appendix A.12 displays the mean of the variable advanced computer use by occupation 2-digits.

Figure 1.1 shows the proportion of workers with an advanced use of computer by agglomeration density. It displays a threshold above and including the category 16, that is an agglomeration with more than 274 inhabitants per km^2. A t-test on the means above and below it (columns 2 and 3 of Table 1.1) shows that the difference is significantly

[30]Second digit for managerial activities.

[31](also 4 dummies: primary education (ISCED 1), ISCED 2 and 3, ISCED 5, ISCED 6 and 7 (university degree))

[32]see https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/510013/VocationalQualificationsNote2016.pdf, visited the 26th of March 2019.

[33]http://www.bbsr.bund.de/BBSR/DE/Raumbeobachtung/Raumabgrenzungen/StadtLandRegionen_Typen/StadtLandRegionen_Typen.html?nn=443270, visited the 14th of January 2017.

Table 1.1: Advanced use of computer at work: descriptive statistics

(a) Germany

Mean (StD) *Observations*	Whole sample (1)	Large cities[b/] (2)	Rest of the country (3)	|t-stat| [c/] (4)
Whole sample	.24	.27	.21	8.2***
	(.43)	(.44)	(.41)	
	18,467	*8,921*	*9,546*	
University degree[a/]	.38	.41	.36	3.7***
	(.49)	(.49)	(.48)	
	4,951	*2,675*	*2,276*	
No university degree[a/]	.19	.21	.17	5.3***
	(.39)	(.40)	(.38)	
	13,516	*6,246*	*7,270*	

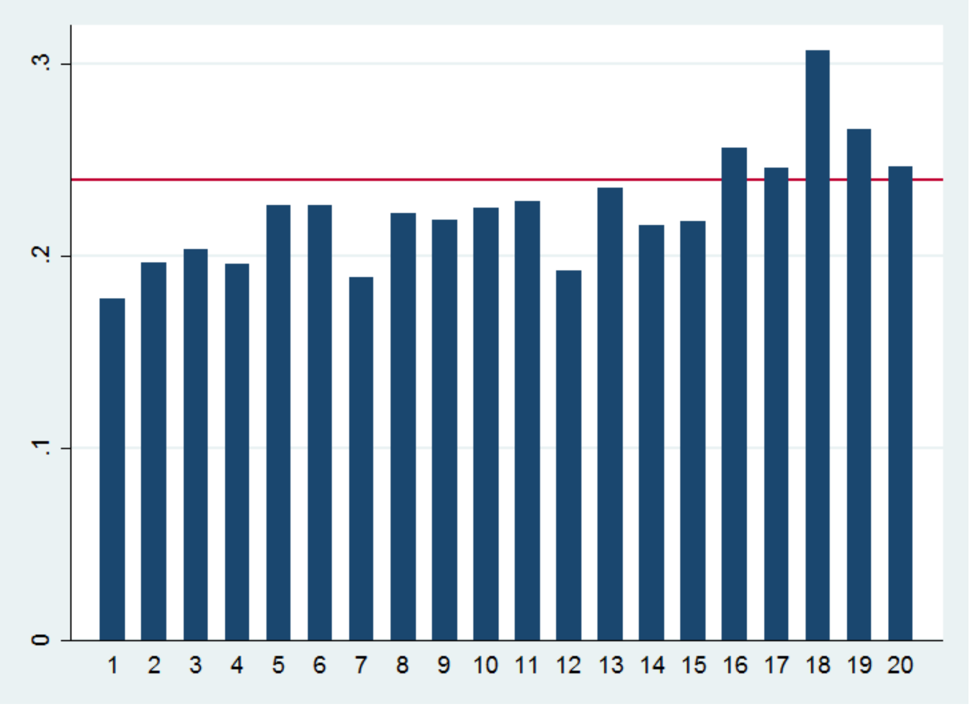
(b) Great Britain

Mean (StD) *Observations*	Whole sample (1)	Large cities[d/] (2)	Rest of the country (3)	|t-stat| [c/] (4)
Whole sample	.20	.23	.19	4.6***
	(.40)	(.42)	(.39)	
	7,010	*2,697*	*4,313*	
University degree[a/]	.36	.40	.33	3.1***
	(.48)	(.49)	(.47)	
	1,689	*725*	*964*	
No university degree[a/]	.16	.17	.15	2.4**
	(.36)	(.38)	(.35)	
	5,321	*1,972*	*3,349*	

[a/] Education required for occupation.
[b/] As defined by to the German administration http://www.bbsr.bund.de/BBSR/DE/Raumbeobachtung/Raumabgrenzungen/StadtLandRegionen_Typen/StadtLandRegionen_Typen.html?nn=443270, visited the 14th of January 2017.
[c/] T-statistic in absolute value rejecting equalities of means in column (2) and (3), assuming unequal variance, with *** p<0.01, ** p<0.05, * p<0.1.
[d/] TTWAs with more than 440′000 inhabitants.
Source: author's computation based on the 2006th waves of the British Sill Survey and the German working population survey.

different from zero. Also, the Spearman correlation between the large cities and having an advanced use of computer at work is .06, where the hypothesis that both variable are independent is rejected at very high significance level.

Figure 1.1: Proportion of workers with an advanced use of computer by category of population density in Germany



Source: author's computation based on the 2006th wave of the German working population survey (18'467 observations).
Notes: the x-axis represents the category of population density (See Appendix A.7 for details). The red line is at 24% of workers with an advanced use of computer and emphasizes the threshold in computer use at the category of integrated labor market 16. Indeed, no category below 16 is above the 24% line, whereas no category above and included it is below that line.
The peak at the category 18 is probably linked to the presence of Bonn in this category. Indeed, as Berlin was chosen to be the new capital city after the fall of the Berlin Wall, Bonn received some compensations in terms of the localization of administration services or research center among other things. Moreover, the agglomerations of the region "Ruhr" are very closed to each other geographically. These agglomerations are spread out between the categories 18 to 20.

Figure 1.2 shows the raw correlation for Great Britain between the proportion of workers with an advanced computer use and city size by agglomeration as I have a continuous measure of size available.[34] The relationship is robust to the exclusion of London which might be considered an outlier, and to the exclusion of agglomerations with less than 20

[34]TTWAs with zero workers with an advanced use of computer and those with half of the population with an advanced use of computer are outliers with less than 20 observations. The former group of TTWAs includes Tunbrigde Wells, Trowbridge and Warminster, Kidderminster, Leek, Louth, Blackpool, Nelson and Colne, Ruthin and Bala, Llangefni and Amlwch, Bervickshire, Hawick, Banff, Dufftown, Dumbarton, Melton Mowbray and Oakham. The latter group of TTWAs includes Andover, Chichester, Crieff.

observations in the sample, where the mean of advanced computer use might be far off the population mean.[35]

Further descriptive statistics are provided in Appendix A.3, which shows the distribution of skill groups across city categories in the sample.

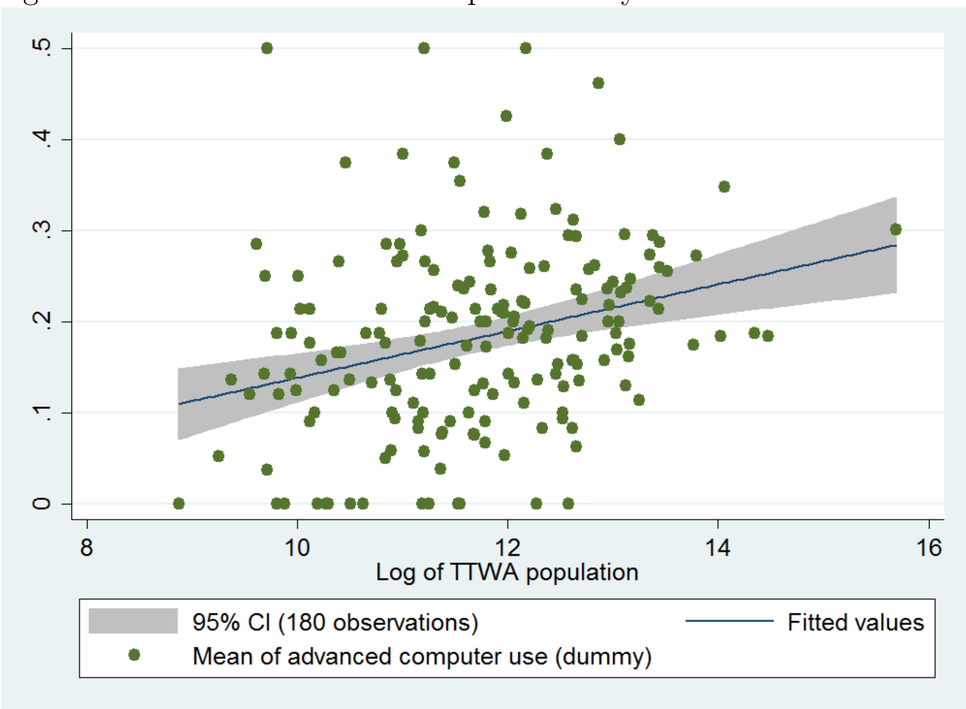
1.3.3 Results advanced computer use at work and city size

Germany

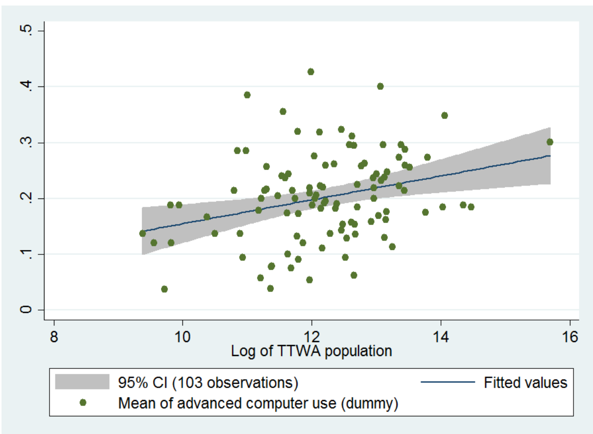
Table 1.2 shows the results of equation (1.12). The main robustness tests consist of excluding rural areas from the samples and I also look at the difference in computer capital across small towns and rural areas (excluding large urban centers), showing that the correlation between city size and computer capital is not driven by a few cities. Moreover, a result consistent for workers with and without a university degree indicates that the correlation between city size and computer capital is not driven by a few occupations. Last, the specific context of Great Britain or Germany is not driving the results.

[35] I also weighted the relationship by population. There is no visual differences in this case compared to the benchmark case (Figure 1.2). I thus do not report the figure in order to save space.

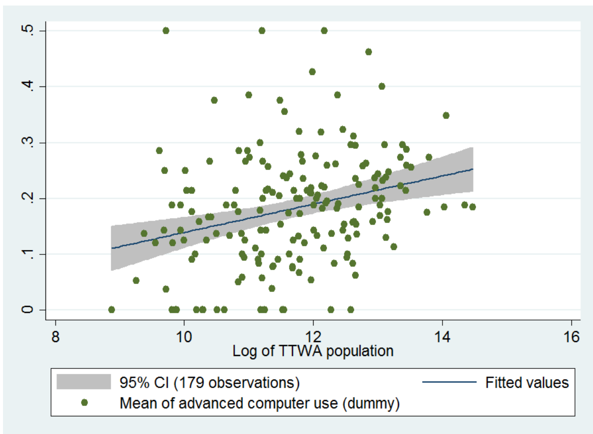
Figure 1.2: Mean of advanced computer use by TTWA in Great Britain



(a) Whole sample



(b) TTWA with more than 20 observations



(c) London excluded

Source: author's computation based on the 2006th wave of the British Sill Survey (7,010 observations).
Notes: the travel to work area (TTWA) represents an integrated labor market based on commuting flows. The slope in (a) is .026 with a standard error of .0060. The slope in (b) is .021 with a standard error of .0058. The slope in (c) is .025 with standard error .0064. All three coefficients are significantly different from zero at the 1% level.
See Table A.1 for the number of observations by log of TTWA population.

Table 1.2: Advanced computer use and city size in Germany (logistic marginal effects at mean workers' characteristics)

Dep. var. Pr(advanced computer) Indep. var.	(1) Whole sample	(2) No rural areas	(3) No large cities	(4) No university[a]	(5) University[a]	(6) Subsample[b]	(7) Whole sample[c]
Large city	.040***	.039***		.032***	.045***	.037***	.014**
	(.0062)	(.0072)		(.0065)	(.0144)	(.0113)	(.0063)
Small towns vs rural			.005				
			(.0083)				
Education[d], age, sex	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Overall score on school leaving certificates[e]	No	No	No	No	No	Yes	No
Occupation and industry fixed effects[f]	No	No	No	No	No	No	Yes
Observations	18,467	14,773	9,546	13,516	4,951	7,342	18,054

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

[a] Education required for occupation.

[b] Overall scores have additional missing values. The benchmark regression of column (1) with this subsample leads to slightly lower estimates: .035*** (.0113), where the subsample overestimate workers in occupation requiring a university degree (27% in the total sample and about 52% in the subsample)

[c] Some observations drop because predicting success or failure perfectly.

[d] 4 education dummies including primary education, secondary 1 and 2, secondary 3 without a university degree and secondary 3 with a university degree.

[e] Very good, good, satisfactory, fair/pass.

[f] 298 occupations (4-digits) and 61 industries (2-digits).

Source: author's computation based on the 2006th wave of the German working population survey.

Notes: Large cities, small towns and rural areas are defined according to the German administration http://www.bbsr.bund.de/BBSR/DE/Raumbeobachtung/Raumabgrenzungen/StadtLandRegionen_Typen/StadtLandRegionen_Typen.html?nn=443270, visited the 14th of January 2017.

Column (1) of Table 1.2 implies that a worker with mean characteristics has a probability of .24 in large cities instead of .20 in the rest of the country to have an advanced computer use at work. Columns (2) and (3) show that the effect takes place exclusively between large cities and small tows, not between small towns and rural areas. Both workers with and without a university degree are concerned by the phenomenon (columns (4) and (5)), where the coefficient is larger for high-skilled workers. Next, column (6) includes overall score on school leaving certificate as an additional measure of skill. This variable is only available for a subsample of 7,342 observations which over represents workers in occupations requiring a university degree (27% in the total sample and about 52% in the present subsample). I thus compare the estimate of .037 to the benchmark estimate (column (1)) of .035 in this subsample, which is very close. Finally, controlling for detailed firm and workers' characteristics (column (7)) reduces the effect by a factor of 3, though it is still significantly different from zero. Appendix A.6 takes the four options of the variable on the use of computer as the dependent variable using a generalized order logit, instead of creating a dummy variable.[36] Also in this specification, workers in large urban centers have on average a more advanced computer use given workers' characteristics. Moreover, excluding large cities also leads to insignificant coefficients.

Great Britain

Regarding Great Britain, Table 1.4 shows a positive relationship between city size and the probability to have an advanced use of computer at work given workers' characteristics. I include column (1) for comparison with the German estimates. It shows that a worker with mean characteristics has a probability of .19 instead of .16 to have an advanced computer use in large cities relative to the rest of the country. The estimate of column (2) implies that a worker with mean characteristics in a city with mean size has a probability of 18.8% instead of 17.4% of having an advanced computer use at work when city size increases by 100%. As the city with mean population has about 270,000 inhabitants (TTWA: e.g. Swansea), an increase of 100% of population is a city with 540,000 inhabitants, e.g. Southampton and Winchester. Columns (3) and (4) displays the main robustnesses, where rural areas and large cities are excluded from the regression. It shows that the complementarity holds in both cases. Also columns (5) and (6) show that both the subsample of workers working in an occupation requiring a university degree or not are concerned by the phenomena, where the coefficient for university workers seems significantly larger. Column (7) includes two additional controls for the skill of workers: the type of school attended as well as highest math qualification with 4,825 non missing observation which over represent workers in occupation requiring a university degree (30% instead of 24% of observations). Population is also significant in this case, where the coefficient is to be compared with the replication of column (1) on the subsample: .0177*** (.0039). Column (8) finally focuses on within detailed occupation variation. It shows that the same job as defined by occupation (4 digits) and industries (1 digit) is done differently in another location.

[36] "No computer"; "Basic use of computer"; "Writing programs for oneself"; "Writing programs for others"

Table 1.3: Advanced computer use and city size in Great Britain (logistic marginal effects at mean workers' characteristics)

Dep. var. Pr(advanced computer)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Indep. var.	Whole sample	Whole sample	No rural areas	No large cities	No university[a]	University[a]	Subsample[b]	Whole sample[c]
Large city	.031***							
	(.0117)							
ln pop		.0140***	.0130***	.0156**	.0072**	.0270***	.0158***	.0078*
		(.0031)	(.0045)	(.0062)	(.0033)	(.0079)	(.0040)	(.0042)
Education[d], age, sex	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Type of school last attended[e]	No	No	No	No	No	No	Yes	No
Highest math qualification	No	No	No	No	No	No	Yes	No
Occupation and industry fixed effects[f]	No	No	No	No	No	No	No	Yes
Observations	7,010	7,010	6,063	4,313	5,321	1,689	4,825	5,784

Std. Err. adjusted for 180 clusters (TTWA)

*** p<0.01, ** p<0.05, * p<0.1

[a] Education required for occupation.

[b] The type of school attended and math qualification have additional missing values. The benchmark regression of column (1) with this subsample leads to slightly higher estimates: .0177*** (.0039), where the subsample overestimate workers in occupation requiring a university degree (30% instead of 24% of observations)

[c] Some observations drop because predicting success or failure perfectly.

[d] 5 categories according to the National Vocational Qualification levels.

[e] Private, grammar public, comprehensive public, secondary modern school, city technology college, other.

[f] 338 occupations (4-digits) and 9 industries (1-digit).

Source: author's computation based on the 2006th wave of the British Sill Survey.

Notes: Ln pop is the log of population by travel to work area which represents an integrated labour market based on commuting flows. Rural areas are defined as TTWA with less than $60,000$ inhabitants and large TTWAs are defined as TTWAs with more than $440,000$ inhabitants.

Last, I take advantage of having a continuous measure of city size available for Great Britain by instrumenting city size with an IVprobit (column (2) of Table 1.4). The estimate of column (2) implies that a worker with mean characteristics has a probability of 19.5% instead of 17.4% of having a advanced computer use at work when city size increases by 100%. Column (1) shows the probit version of column (1) in Table 1.3, as instrumenting with a logistic regression is not feasible.

Table 1.4: Advanced computer use and city size in Great Britain (probit marginal effects at mean workers' characteristics): instrumenting city size

Dep. var. Pr(advanced use of computer) Indep. var.	(1) Probit	(2) IVProbit$^{a/}$
ln pop	.0148*** (.0034)	.0249*** (.0040)
Worker's characteristics$^{b/}$	Yes	Yes
Occupation and industry fixed effects	No	No
Observations	7,010	7,010
MOP Effective F statistics$^{c/}$		640

Std. Err. adjusted for 180 clusters (TTWA)

*** p<0.01, ** p<0.05, * p<0.1

$^{a/}$ First stage: $ln(population) = .8^{*} + 1.02^{***}ln(population1851)\ -.2women - .002^{**}age + .7^{**}educ1 + .1^{***}educ2 - .005educ3 + .05educ4$.

$^{b/}$ Education (5 categories according to the National Vocational Qualification levels), age and sex.

$^{c/}$ Montiel Olea-Pflueger robust weak instrument test (2013): effective first-stage F statistic. A critical value of 37.4 entails a 5% of worse case bias.

Source: author's computation based on the 2006th wave of the British Sill Survey.

Notes: Ln pop is the log of population by travel to work area (TTWA) which represents an integrated labour market based on commuting flows.

1.3.4 Further robustness

Using additional proxies of computer capital including "having a computer at work" and "working with a computer as main work equipment" as proxy for computer capital confirm the positive relationship between computer capital and city size. There is nevertheless no significant differences across city sizes in the probability to work with a computer at work for workers in occupation requiring a university degree. The second difference with the benchmark case is that the relationship between computer capital and city size remains significant when large cities are excluded from the sample (see Appendix A.8).

Moreover, the relationship between the probability to have an advanced use of computer at work and city size is significant in all industries in Germany except for the public sector, where the largest effect is found in manufacturing. The British public sector also shows no effect, and the largest estimate is from the finance industries. The service sector is thus not the only one impacted by the concentrating force of computerization (see Appendix A.8).

1.3.5 Comparison of German and GB datasets

Qualitatively, the estimates in the German and the Great British datasets are consistent. The first main difference is that excluding large cities drives the coefficient to zero in the German case, but not in Great Britain (Table 1.2 column (3) and Table 1.3 column (4)). The second main difference is that the coefficient for workers in occupations requiring a university degree is significantly larger than the coefficient for other workers in the British case, but not in Germany (Table 1.2 columns (5) and (6) and Table 1.3 columns (3) and (4)).

I now turn to a comparison of the magnitude of the estimates of the benchmark cases in Tables 1.2 and 1.4 as well as Tables A.6 and A.7. Column (1) shows that the increase in the probability to have an advanced use of computer at work in large cities relative to the rest of the country is slightly larger in Germany than in Great Britain (.040 versus .031 percentage points). The same is true regarding having a computer at work even though the number are very similar (.035 versus .033 percentage points). Nevertheless, my definition of large cities in Great Britain rests on no official definition as is the case for Germany. Appendix A.9 thus compares the German and British coefficients based on an increase in city size and city rank, which gives similar results.

1.4 Mechanisms underlying the empirical relationship between computer capital and city size

The purpose of this section is to explore the mechanisms behind the complementarity of city size and computer capital as highlighted in Section 1.3. I will discuss them in light of the German BIBB dataset as most of the variable I am using have no equivalent in the British Skill Survey, focusing on the advanced use of computer which is my benchmark proxy for computer capital and in relation with the economic literature.

1.4.1 Complex tasks and unobserved skill

I first consider the mechanism I expose in the model of this paper. In it, the complementarity of city size and computer capital relies on complex tasks. As city population disproportionately increases the productivity of complex tasks relative to simple ones, firms have a stronger incentives to invest in the capital complementary with complex tasks the larger the city, given the skill of workers. This subsection aims at documenting if complex tasks in the job and computer capital allocation go hand in hand, given the skill of workers.

Data and empirical strategy

I this subsection I need a measure for the tasks' content of the job of workers. First of all, I use detailed occupations. Indeed, detailed occupations of workers are also constructed on job's tasks' content.[37] Second, I use direct measures of tasks as there is a within occupation variation of tasks across time and location (e.g. Spitz-Oener (2006), Kok (2014), Appendix A.10). I divide non-routine cognitive tasks between non-rountine analytical ("Organizing, making plans, working out operation", "Research, development", "Gathering information, investigating, documenting") and non-rountine interactive tasks ("Purchasing, selling", "Promoting, marketing, public relations", "Teaching, training", "Consulting, advising").[38] In the BIBB survey, workers indicate that they perform these tasks "never", "sometimes" or "often". I code "never" 0, "sometimes" .5 and "often" as 1. I then calculate the average for each worker of the three analytical variables and the four interactive variables.

Alternatively, I use as a proxy for the share of time allocated to complex tasks, the share of time working on computers. Indeed, in the model, a worker performs complex tasks with the help of computer capital. Nevertheless, work on computer might also be routine (E.g. bookkeeping). Figure 1.3 though shows that highest educated workers overall tend to work more frequently on computer.[39] Table 1.5 shows some descriptive statistics for the two indexes, as well as for the share of time working on computer. Appendix A.10 regresses these indexes on the large city dummy given workers' characteristics.

The main problem of Section 1.3 though remains. I do not have a comprehensive measure of skill available and complex tasks might well capture unobserved skill. The variation of complex tasks within occupation across years was indeed used to proxy change in skill within occupations (Spitz-Oener (2006)). As such, I unfortunately cannot distinguish between complex tasks and unobserved skill.

I run a version of equation (1.12) controlling in turns and jointly for workers' characteristics, occupations and the individual complex tasks' index. If $\gamma_1 = 0$ in equation (1.13) I conclude that the effect of city size on the probability to have an advanced use at work acts through the workers' tasks' content:

$$P(c_i = 1|x_i) = G(\gamma_0 + \gamma_1 LargeCity_j + \gamma_2 Ind_i + \gamma_3 Complex_i) \tag{1.13}$$

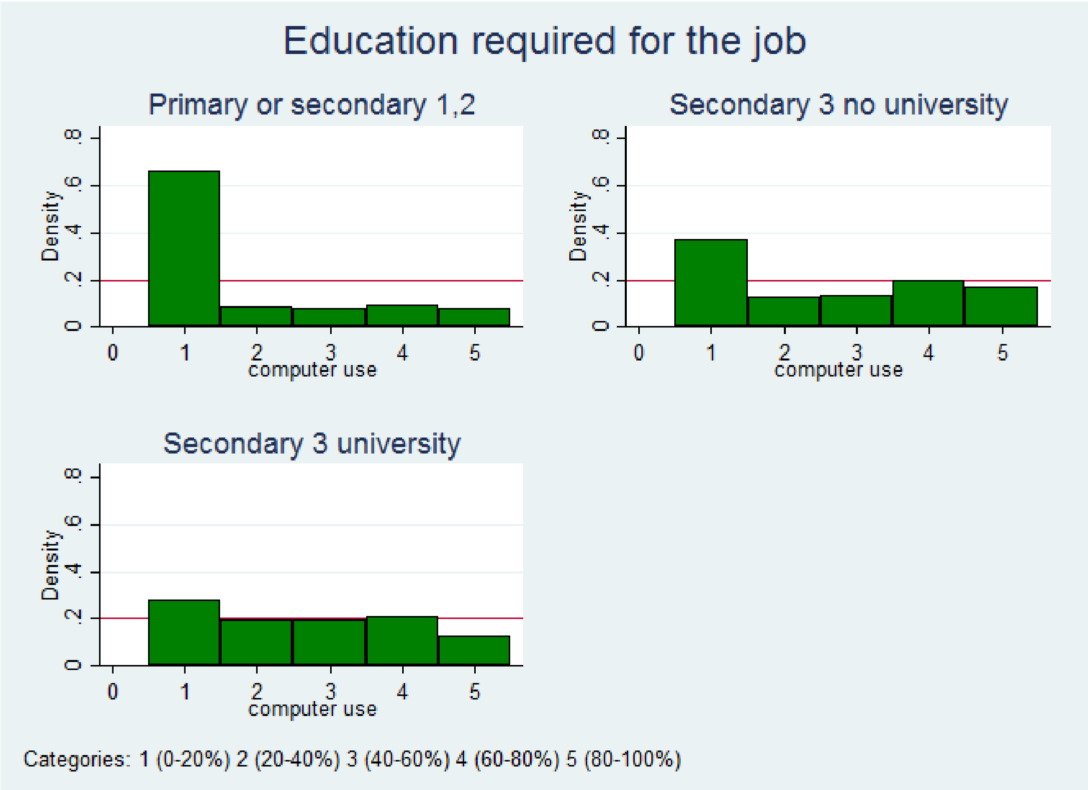
[37] The lowest unit in the ISCO88 classification of occupations used in the BIBB survey is defined as a "set of tasks or duties designed to be executed by one person" (Elias and Birch (1994), p.1.)

[38] See e.g. Spitz-Oener (2006), Table 1, p.243.

[39] Two additional caveats of this proxy are first that complex tasks might be performed without computer, even though in complementarity with it to produce a complex output. Second, the percentage of time working on computer might also capture non-work activities such as playing Tetris or surfing on Youtube.

where c_i is a binary variable which equals 1 if worker i has an advanced use of computer at work, $LargeCity_j$ is a dummy variable indicating that city j of worker i is a city with more than 274 inhabitants per km^2, controls Ind_i include worker's characteristics - 4 education dummies, age and sex, 298 occupations (4-digits) and $Complex_i$ is the complex tasks' index. $G(.)$ is the logistic cumulative distribution function.

Figure 1.3: Time spent on computer by education required for the job in Germany



Source: author's computation based on the 2006th wave of the German working population survey with sampling weight.

Notes: the first group corresponds to jobs requiring primary education and the first and second stages of secondary education or apprenticeship, the second group to jobs requiring about 4 year of education after the age of 17 or 18 not equivalent to a university degree, the last one to jobs requiring a university degree. I merge primary and the first and second stages of secondary education or apprenticeship due to the low number of observations across computer use for primary education. The red lines at a density of 2 help visualize the difference in the density of % computer use at work by education required for the job.

Table 1.5: Measure of complex tasks: descriptive statistics

(a) Analytical tasks index

Mean (StD) *Observations*	Whole sample (1)	Large cities[b/] (2)	Rest of the country (3)	|t-stat| [c/] (4)
Whole sample	.48	.50	.46	8.3***
	(.28)	(.28)	(.28)	
	18,467	*8,921*	*9,546*	
University degree[a/]	.63	.63	.63	0.3
	(.24)	(.24)	(.25)	
	5,330	*2,908*	*2,422*	
No university degree	.42	.43	.41	5.4***
	(.27)	(.27)	(.27)	
	13,137	*6,013*	*7,124*	

(b) Interactive tasks index

Mean (StD)	Whole sample (1)	Large cities[b/] (2)	Rest of the country (3)	|t-stat| [c/] (4)
Whole sample	.52	.53	.50	7.1***
	(.26)	(.25)	(.26)	
University degree[a/]	.62	.62	.62	1.4
	(.20)	(.20)	(.20)	
No university degree	.47	.49	.46	5.6***
	(.27)	(.26)	(.27)	

(c) % working time with a computer

Mean (StD)	Whole sample (1)	Large cities[b/] (2)	Rest of the country (3)	|t-stat| [c/] (4)
Whole sample	.39	.43	.36	14.1***
	(.34)	(.34)	(.33)	
University degree[a/]	.48	.51	.45	7.6***
	(.30)	(.30)	(.30)	
No university degree	.36	.39	.33	10.1***
	(.35)	(.35)	(.34)	

[a/] Defined according to formal education.
[b/] As defined by to the German administration http://www.bbsr.bund.de/BBSR/DE/Raumbeobachtung/Raumabgrenzungen/StadtLandRegionen _Typen/StadtLandRegionen_Typen.html?nn=443270, visited the 14th of January 2017.
[c/] T-statistic in absolute value rejecting equalities of means in column (2) and (3), assuming unequal variance, with *** p<0.01, ** p<0.05, * p<0.1.
Source: author's computation based on the 2006th wave of the German working population survey.

Results

Table 1.6: Advanced computer use and city size in Germany (logistic marginal effects at mean workers' characteristics): tasks

Dep.var. Pr(advanced computer) Indep. var.	(1)	(2)$^{a/}$	(3)	(4)$^{a/}$
Large city	.040***	.014**	.012**	.008
	(.0062)	(.0063)	(.0057)	(.0060)
Worker's characteristics$^{b/}$	Yes	Yes	Yes	Yes
Occupation fixed effects 4-digits	No	Yes	No	Yes
Analytical tasks' index	No	No	Yes	Yes
Interactive tasks' index	No	No	Yes	Yes
% working time on computer	No	No	Yes	Yes
Observations	18,467	18,056	18,467	18,056

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

$^{a/}$ Some observations drop because predicting success or failure perfectly.

$^{b/}$ Age, sex and 4 education dummies including primary education, secondary 1 and 2, secondary 3 without a university degree and secondary 3 with a university degree.

Source: author's computation based on the 2006th wave of the German working population survey.

Notes: Large cities are defined according to the German administration http://www.bbsr.bund.de/BBSR/DE/Raumbeobachtung/Raumabgrenzungen /StadtLandRegionen_Typen/StadtLandRegionen_Typen.html?nn=443270, visited the 14th of January 2017.

Column (1) of Table 1.6 reproduces the benchmark estimate of Table 1.2. Column (2) controls for 4 digits occupations, which partly captures tasks and skill of workers. It shows that differences in the sorting of occupations between large cities and the rest of the country explains about 2/3 of the relationship between the probability to have an advanced use of computer and the large city dummy. Column (3) focuses on the direct measures of complex tasks, which can thus vary within occupation across city size. The positive correlation between the probability to have an advanced use of computer and large cities remains significant. It means that either my direct task's measures do not fully capture complex tasks of workers, or that there exists a within-tasks variation in the probability to have an advanced use of computer at work. Last, in column (4), I control simultaneously for occupation and my direct measures of tasks. In this specification, my coefficient of interest becomes insignificant. I nevertheless cannot be sure that it is the tasks perform by the workers at work or unobserved skill correlated with tasks and occupation which explains the variation in computer capital across city sizes. Also, there are many controls, which is maybe too much for the model. Comparing columns (2) and (4), at least half of the within-occupation variation in the probability of having an advanced use of computer across city sizes is explained by the within-occupation variation of tasks' content across city size.

1.4.2 Labor capital complementarity in the presence of agglomeration economies

The second mechanism I consider is labor capital complementarity in the presence of agglomeration economies. A simple model with these features would lead to the prediction that capital's allocation increases with city size as labor is more productive in large city, for any type of capital. In the context of my model, this mechanism translates in the optimal level of both complex and simple capital through larger local GDP Y_A due to agglomeration economies (see equations (A.17) and A.18). Is the relationship highlighted in Section 1.3 a general capital story or is it specific to computer capital?

More generally, how do the 14 equipment types I have in the data relate to the two representative equipment types of my model (simple and complex)? Are there other types of capital which can be considered "complex" in the sense of the model? (assumption 1 part (ii) in Section 1.2.1). What types of capital are "simple" in the sense of the model? What are the dimensions in which the equipment types presented in this section do not fit the model of this chapter?

In this subsection, I look at the behavior of other types of capital with city size, given workers' characteristics. The measure I have available to proxy other types of capital comes from the German BIBB survey. It states among 14 equipment types which one is the main work equipment. The empirical model used is the same than in Section A.8.1. Table 1.7 shows the relationship between the probability to have the considered equipment as main work equipment and the large city dummy by increasing order of association with education. Note that workers reporting that they have no main work equipment is negatively related with the large city dummy given workers' characteristics. Although the coefficient is only significant at the 10% level, it is an indication that equipment in general become more economically profitable as city size increases.

Table 1.7: Main work equipment and city size in Germany (logistic marginal effects at mean workers' characteristics)

Dep. Var. (d.v.) Pr(main equipment)	Large city dummy (Robust SE)	Mean of d.v.	Cramer's V d.v. & educ *sig Chi2*	
Tool	-.012***	.08	-.164	
	(.003)		*.000*	
Transportation	-.007**	.08	-.102	
	(.003)		*.000*	
Machine	-.013***	.04	-.100	simple capital
	(.002)		*.000*	
Clothing	.001	.01	-.048	
	(.001)		*.000*	
Instrument	-.003	.03	-.035	
	(.003)		*.000*	
Operating materials	-.000	.002	-.026	
	(.000)		*.000*	
Equipment[a/]	-.001	.002	-.026	
	(.001)		*.000*	
Manufacturing plant [b/]	-.000	.002	-.022	
	(.000)		*.003*	
No work equipment	-.004*	.03	-.019	
	(.002)		*.010*	
Resource to ease a work	-.002	.02	-.005	
	(.002)		*.492*	
Device	.010**	.09	.003	
	(.004)		*.724*	
Classification not possible [c/]	-.026***	.15	.028	
	(.005)		*.000*	
Furniture [d/]	-.002	.01	.048	complex capital
	(.001)		*.000*	
Computer	.074***	.45	.189	
	(.007)		*.000*	

*** p<0.01, ** p<0.05, * p<0.1

[a/] Equipment, such as hand-controlled or automatic machines (in German: Vorrichtung).

[b/] Technical equipment such as manufacturing plant (in German: Anlage)

[c/] The classification was conducted based on an open question, unfortunately, I do not have access to it.

[d/] Facility, furniture, such as fitness equipment or medical furniture (in German: Einrichtung).

Notes: 18,499 observations, controlling for worker's education (4 dummies), age and sex

The Cramer's V can be negative because both variables are dummies (main work equipment and education - having a university degree).

Source: author's computation based on the 2006th wave of the German working population survey

Work equipment negatively related to education

The first group of equipment types negatively correlates with education and are candidates for being simple capital as defined in this chapter ("simple capital"). They are the first seven equipment types in Table 1.7. According to the model in Section 1.2, their investment per worker can be either increasing or decreasing in city size. Tool, transportation and machine as main work equipment negatively and significantly correlate with the city dummy, given workers' characteristics. However, once education is measured with the first digit of the occupation classification, the city dummy is insignificant for both tool and transportation. In this first group, there is no equipment positively related to urban centers given workers' characteristics.

Table 1.8 offers details for machine as main work equipment. The conditional probability for a worker with mean characteristics to work with it as main equipment is 2.4% in urban centers and 3.6% in the rest of the country. The same figures for workers without a university degree are 3.8% in urban centers and 5.8% in the rest of the country. The coefficient barely changes when rural areas are excluded from the sample and insignificant when comparing small towns with rural areas. The relationship is entirely driven by workers without a university degree, which represent 93 percent of workers having machines as their main work equipment. The relationship is barely significant when controlling for detailed occupations and industries, one fourth of the observations, though drop in this case because of predicting success or failure perfectly. The main departure from simple capital in my model is that machines are usually large equipment so that their investment cost should depend on rent prices. As such they are more expansive in large cities. In Section 1.2 though, I assume that the cost of capital is independent from the cost of housing. Section 1.4.3 discusses difference in the cost of equipment as a function of rent prices as an alternative mechanism underlying the comparative advantage of computer capital with city size relative to other equipment types.

Table 1.8: Maschines as main work equipment and city size in Germany (logistic marginal effects at mean workers' characteristics)

Dep.var. Pr(Maschine main equip.) Indep. var.	(1) Whole sample	(2) No rural	(3) No large cities	(4) No university[a/]	(5) University [a/]	(6) Whole sample[b/]
Large city	-.013***	-.012***		-.021***	-.001	-.002*
	(.0022)	(.0023)		(.0037)	(.0026)	(.0011)
Small towns vs rural			-.000			
			(.0036)			
Worker's characteristics[c/]	Yes	Yes	Yes	Yes	Yes	No
Occupation and industry fixed effects[d/]	No	No	No	No	No	Yes
Observations	18,499	14,800	9,557	13,161	5,338	14,110

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

[a/] According to formal education.

[b/] Some observations drop because predicting success or failure perfectly.

[c/] Age, sex and 4 education dummies including primary education, secondary 1 and 2, secondary 3 without a university degree and secondary 3 with a university degree.

[d/] 298 occupations (4-digits) and 61 industries (2-digits).

Source: author's computation based on the 2006th wave of the German working population survey.

Notes: Large cities, small towns and rural areas are defined according to the German administration, http://www.bbsr.bund.de /BBSR/DE/Raumbeobachtung/Raumabgrenzungen/StadtLandRegionen_Typen/StadtLandRegionen_Typen.html?nn=443270, visited the 14th of January 2017.

Work equipment unrelated to education

The second group of equipment types is unrelated to education at the .1% significance level (four equipment types).[40] Most of them are unrelated with large cities as main work equipment. However, device as main work equipment is positively related to large cities, where the probability to have it as main work equipment is 9.5% in large cities compared to 8.5% in the rest of the country. Table 1.9 shows that the magnitude of the effect is untouched when rural areas are excluded from the sample, that no correlation is highlighted when comparing small towns from rural areas and that the correlation is driving by workers without a university degree. Last, the magnitude of the effect is barely reduced when controlling for detailed occupations and industries.

[40]This mere fact makes it at odd with my model, with two equipment types and a comparative advantage for high-skilled workers in complex computer-tasks

Table 1.9: Device as main work equipment and city size in Germany (logistic marginal effects at mean workers' characteristics)

Dep. var. Pr(Device main equip.) Indep. var.	(1) Whole sample	(2) No rural	(3) No large cities	(4) No university[a/]	(5) University[a/]	(6) Whole sample[b/]
Large city	.010**	.010**		.012**	.009	.008**
	(.0042)	(.0049)		(.0050)	(.0079)	(.0036)
Small towns vs rural			.001			
			(.0059)			
Worker's characteristics[c/]	Yes	Yes	Yes	Yes	Yes	No
Occupation and industry fixed effects[d/]	No	No	No	No	No	Yes
Observations	18,499	14,800	9,557	13,161	5,338	18,343

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
[a/] According to formal education.
[b/] Some observations drop because predicting success or failure perfectly.
[c/] Age, sex and 4 education dummies including primary education, secondary 1 and 2, secondary 3 without a university degree and secondary 3 with a university degree.
[d/] 298 occupations (4-digits) and 61 industries (2-digits).
Source: author's computation based on the 2006th wave of the German working population survey.
Notes: Large cities, small towns and rural areas are defined according to the German administration, http://www.bbsr.bund.de /BBSR/DE/Raumbeobachtung/Raumabgrenzungen/StadtLandRegionen_Typen/StadtLandRegionen_Typen.html?nn=443270, visited the 14th of January 2017.

Work equipment positively related to education

The third group of equipment types is positively related to education and thus candidates to fit the complex capital ideal type of the model ("complex capital?" in Table 1.7, three last equipment types). Among the three work equipment types in this group, only computer capital is positively correlated with large urban centers. Table 1.10 offers details on the relation between large urban centers and not classified main work equipment. The conditional probability for a worker with mean characteristics to work with equipment not classifiable is 13.3% in urban centers and 15.9% in the rest of the country. The conditional probability for a worker with a university degree with mean characteristics to work with an equipment not classifiable is 12.9% in urban centers and 18.9% in the rest of the country. Next, Table 1.11 details the results for furniture as main work equipment. Overall, the positive correlation of the equipment with education is not systematically related to a positive relationship with city size. It gives a hint that unobserved skill is not fully explaining the relationship between computer capital and city size. The correlation with education is though stronger for computer capital than other equipment.

Table 1.10: Equipment not classified and city size in Germany (logistic marginal effects at mean workers' characteristics)

Dep. var. Pr(Equipment not classified) Indep. var.	(1) Whole sample	(2) No rural	(3) No large cities	(4) No university[a/]	(5) University[a/]	(6) Whole sample[b/]
Large city	-.026***	-.022***		-.009	-.060***	-.008*
	(.0052)	(.0058)		(.0061)	(.0097)	(.0041)
Small towns vs rural			-.007			
			(.0076)			
Worker's characteristics[c/]	Yes	Yes	Yes	Yes	Yes	No
Occupation and industry fixed effects[d/]	No	No	No	No	No	Yes
Observations	18,499	14,800	9,557	13,161	5,338	18,409

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

[a/] According to formal education.

[b/] Some observations drop because predicting success or failure perfectly.

[c/] Age, sex and 4 education dummies including primary education, secondary 1 and 2, secondary 3 without a university degree and secondary 3 with a university degree.

[d/] 298 occupations (4-digits) and 61 industries (2-digits).

Source: author's computation based on the 2006th wave of the German working population survey.

Notes: Large cities, small towns and rural areas are defined according to the German administration, http://www.bbsr.bund.de /BBSR/DE/Raumbeobachtung/Raumabgrenzungen/StadtLandRegionen_Typen/StadtLandRegionen_Typen.html?nn=443270, visited the 14th of January 2017.

Table 1.11: Furniture as main work equipment and city size in Germany (logistic marginal effects at mean workers' characteristics)

Dep. var. Pr(Furniture main equip.) Indep. var.	(1) Whole sample	(2) No rural	(3) No large cities	(4) No university[a/]	(5) University[a/]	(6) Whole sample[b/]
Large city	-.002	-.001		.000	-.007*	.000
	(.0014)	(.0016)		(.0015)	(.0033)	(.)
Small towns vs rural			-.000			
			(.0017)			
Worker's characteristics[c/]	Yes	Yes	Yes	Yes	Yes	No
Occupation and industry fixed effects[d/]	No	No	No	No	No	Yes
Observations	18,499	14,800	9,557	13,161	5,338	

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

[a/] According to formal education.

[b/] Some observations drop because predicting success or failure perfectly.

[c/] Age, sex and 4 education dummies including primary education, secondary 1 and 2, secondary 3 without a university degree and secondary 3 with a university degree.

[d/] 298 occupations (4-digits) and 61 industries (2-digits).

Source: author's computation based on the 2006th wave of the German working population survey.

Notes: Large cities, small towns and rural areas are defined according to the German administration, http://www.bbsr.bund.de /BBSR/DE/Raumbeobachtung/Raumabgrenzungen/StadtLandRegionen_Typen/StadtLandRegionen_Typen.html?nn=443270, visited the 14th of January 2017.

1.4.3 Competing mechanisms

Difference in the size of equipment and rents

The third mechanism I consider is that the higher cost of housing in large cities disadvantages large equipment types relative to smaller ones. Indeed, firms also have to pay rent to host equipment. In my model in Section 1.2, I though assume that the cost of capital is independent from location. As computerized equipment is relatively small, this might explain its prevalence in large cities relative to larger equipment such as machines (see Table 1.7). Because the manufacturing sector relies on larger equipment, the increase in rents following an increasing attractiveness of large cities might lead it to relocate factories in smaller towns.

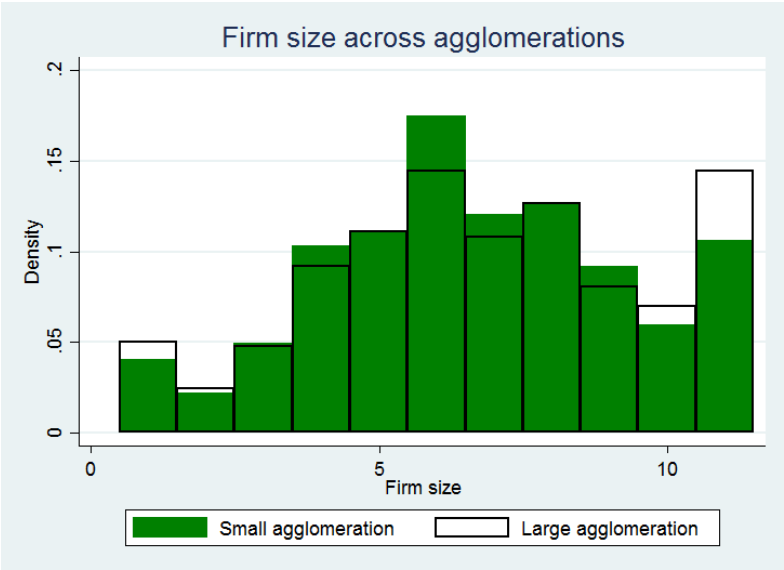
Infrastructures

The fourth mechanism I consider relates to infrastructures. Because of a larger market and a more favorable geography (denoted A in the model of Section 1.2), infrastructures tend to be more developed in large cities. In particular, access to public infrastructures such as broadband coverage differs across city sizes which increases the benefits of computer use (see Salemink et al. (2017) for a recent literature review on the urban-rural "digital divide" in advanced western countries). Since rural areas are likely to be the most disadvantaged in this respect, I show that the relationship between computer capital and city size is robust to excluding rural areas. Column (2) in Tables 1.2, 1.3, A.3, A.6, A.7 and A.8 constantly shows that this is the case. The decrease in the coefficient of large cities is very small for Germany and only slightly larger for Great Britain. I though cannot exclude that relevant differences in infrastructures between large cities and the rest of the country are at the core of the relationship. Indeed, column (3) of the Tables mentioned above shows that the coefficient of interest reduces considerable when large urban centers are excluded for Germany and become insignificant when computer capital is proxied by having an advanced use of computer at work. This is not the case for Great Britain where the magnitude of the coefficient slightly increases when large cities are excluded from the sample.

Internal returns to scale

The fifth mechanism I consider is internal returns to scale. Firm size is also related to population density as well as with computer capital. Indeed, as Figure 1.4 shows, firm size distribution is more dispersed in dense areas. Moreover, larger firms tend to invest more in computer capital as they can spread the cost of equipment between a larger pool of employees due to internal return to scale. Table 1.12 presents the results of equation (1.12) additionally controlling for firm size. It shows that the city-computer capital complementarity is very similar to the numbers in Section 1.3.3. I thus exclude internal returns to scale as the main driver of the complementarity.

Figure 1.4: Firm size and city density in Germany



Source: author's computation based on the 2006th wave of the German working population survey

Table 1.12: City and computer in Germany controlling for firm size (logistic marginal effects at mean workers' characteristics)

Indep. var. \Dep. var.	(1) Pr(computer main equip.)	(2) Pr(computer)	(3) Pr(advanced computer)
Large city	.066***	.027***	.034***
	(.0076)	(.0044)	(.0062)
Workers' characteristics[a/]	Yes	Yes	Yes
Firm size dummies	Yes	Yes	Yes
Observations	18,467	18,467	18,467
Mean dep. var.	.45	.83	.24
Mean dep. var. educ==1	.60	.97	.36

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

[a/] Age, sex and 4 education dummies including primary education, secondary 1 and 2, secondary 3 without a university degree and secondary 3 with a university degree.

Source: author's computation based on the 2006th wave of the German working population survey

Notes: Large cities, small towns and rural areas are defined according to the German administration, http://www.bbsr.bund.de/BBSR/DE/Raumbeobachtung/Raumabgrenzungen/StadtLandRegionen_Typen/StadtLandRegionen_Typen.html?nn=443270, visited the 14th of January 2017.

Specialization on core tasks

The sixth possible mechanism I consider is specialization on core tasks. Kok (2014) shows that city size is associated with a stronger specialization on core tasks given occupation and industry. Complex tasks on computer might be prevalent in large cities because they are the core tasks of several occupations. A large proportion of occupations concerned by the phenomenon would weaken this hypothesis. Columns (4) and (5) of Tables 1.2, 1.3, A.3, A.6, A.7 and A.8, split the sample between occupations requiring a university degree versus those who do not require it. They show that the complementarity holds for both groups, except for the proxy of computer capital “working with a computer”, where the relationship is insignificant for workers with a university degree.

Next, I measure specialization by the prevalence of computer capital in an occupation. Indeed, some technical professions do not require a university degree but are intensive in computer capital. I measure three level of computer capital: 1. Working with a computer without writing programs or using macro; 2. Workers writing programs or using macro, and only as users; 3. Workers using computers beyond user activities. Indeed, each of these computer activities require more computer power and as such more investment in computer capital. I take the average of the computer capital proxy taking values one, two and three by 4-digit level occupations. I then defined an occupation as being intensive in computer capital if the average of the computer capital proxy is above the median occupation value and else as non intensive in computer capital. Table 1.13 below displays the relationship between city size and computer capital for the two subsamples.

The regression displayed in Table 1.13 uses the 4 possible values available in the survey for computer use of worker i c_i instead of creating a dummy. Because the values of the variables can be ordered (1.Not working with a computer; 2. Working with a computer without writing programs or using macro; 3. Workers writing programs or using macro, and only as users; 4. Workers using computers beyond user activities.), I use a generalized ordered logistic model.[41]

$$P(c_i > j|x_i) = G(\gamma_{0,j} + \gamma_1 Pop_j + \gamma_2 Ind_i), \quad j = 1, 2, 3 \tag{1.14}$$

where the variables are the same than in equation (1.12) and G is the logistic cumulative distribution function.

[41] See Williams (2006).

Table 1.13: Computer capital and city size in Germany by occupation intensity in computer capital (generalized ordered logistic marginal effects at mean workers' characteristics)

Indep. var.	Dep. var.[a/]	(1) Occupations intensive in computer capital	(2) Occupations not intensive in computer capital
Large city	1	-.008***	-.029**
		(.0025)	(.0136)
	2	-.029***	.024*
		(.0084)	(.0136)
	3	.017***	.004
		(.0065)	(.0049)
	4	.020***	.001
		(.0062)	(.0041)
Worker's characteristics[b/]		Yes	Yes
Occupation and industry		No	No
Observations		12,688	5,779

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
[a/] The dependent variable takes the following values: 1.Not working with a computer; 2. Working with a computer without writing programs or using macro; 3. Workers writing programs or using macro, and only as users; 4. Workers using computers beyond user activities.
[b/] Age, sex and 4 education dummies including primary education, secondary 1 and 2, secondary 3 without a university degree and secondary 3 with a university degree.
Source: author's computation based on the 2006th wave of the German working population survey.
Notes: Large cities are defined according to the German administration http://www.bbsr.bund.de/BBSR/DE/Raumbeobachtung/Raumabgrenzungen/ StadtLandRegionen_Typen/ StadtLandRegionen_Typen.html?nn=443270, visited the 14th of January 2017.

Table 1.13 shows that the positive relationship between city size and computer capital is present in both occupations intensive in computerized equipment (column (1)) and not intensive in it (column (2)). The significance level are though lower for occupations not intensive in computerized equipment and the additional computer capital in large cities relative to the rest of the country translates in working with a computer use without writing programs or using macro. For occupation intensive in computer capital the probability that a worker has an advanced computer use increases whereas the probability of a simple computer use decreases. Overall, I reject the hypothesis of specialization as the main driver of the city size - computer capital complementarity. Indeed, it would require the signs in column (2) to have opposite direction where the probability to work with a computer be *decreasing* in city size.

Computerization and services

Next, computerization also allows to perform given tasks at distance. Some occupations that were previously fully non-tradable can now benefit from agglomeration economies in large cities. For example, the health sector can now have highly productive units in large cities providing part of diagnostic all over the country. In this respect, computerization is also a centralizing force.[42] This mechanism would though also translate in a difference of tasks content of jobs across occupation. Some tasks can become tradable with computerization, while others remain non-tradable. The increasing number of tradable tasks result in an increasing sorting of them across city size, where more complex tasks sort to larger cities.

Globalization

The last alternative mechanism I consider is globalization. When China opens to trade, some tasks are offshored there and it tends to be the routine ones (see e.g. Goos et al. (2014)). A given level of technology is necessary for this to happen as I highlighted in the previous mechanism, but then, for a given price of computer, the opening to trade of a country leads to a increasing proportion of complex tasks in the job in high-wage countries. As a result the attractiveness of large cities increases as well as the amount of computerized equipment allocated to workers, based on the mechanism I highlight in Section 1.2.

As all mechanisms highlighted in this subsection can disproportionately benefit high-skilled workers, they can explain their increasing sorting in large cities. To summarize, the main mechanisms beside complex tasks or unobserved skill I retain as the main driver of the computerized equipment - city size complementarity are differences in the size of equipment and infrastructures specific to large cities.

1.5 Evidence on wages, city size, computer capital and education

In this section, I take the empirical wage implication of relaxing the perfect mobility assumption. I regress - as a correlation - the log of wages $log(w_i)$ of worker i on the triple interaction of the log of population of the agglomeration j in which the worker lives $log(Pop_j)$, her skill, measured by having a university degree Uni_i and computer capital allocated to her, proxied by having an advanced use of computer c_i, where I expect the coefficient of interest $\eta_7 > 0$ (see Section 1.3.1 for a discussion of the limitations of my proxies):

[42] See Combes and Gobillon (2015) for a review of the literature on the impact of a decrease in transportation costs in the context of the New Economic Geography (Krugman (1991)).

$$\ln(w_i) = \eta_0 + \eta_1 \ln(Pop_j) + \eta_2 Uni_i + \eta_3 c_i + \eta_4 \ln(Pop_j) Uni_i + \eta_5 log(Pop_j) c_i + \eta_6 c_i Uni_i + \eta_7 \ln(Pop_j) c_i Uni_i + \epsilon_i \quad (1.15)$$

$\eta_7 > 0$ in equation (1.15) correctly identifies the difference in the productivity of computer capital for high-skilled relative to low-skilled with city size provided that workers' probability to have an advanced use of computer as a function of city size according to unobserved skill is the same for workers with and without a university degree, as I compare them. Whereas I also instrument city size with long-lagged population, I lack an instrument for computer capital.

I focus on individual wages as the dependent variable from the British Skill Survey. I construct it by dividing the "weekly gross pay" by the "usual number of hours worked per week including paid or unpaid overtime". I keep workers from England, Scotland and Wales working full-time. Taking into account missing observations for hourly wage as well as unrealistic observations[43], this sample contains $4,117$ observations. The distribution of the five education categories is remarkably similar in this subsample relative to the whole sample of the BSS. However, documenting wage is not the first purpose of the survey, and asking employees about their wage might be subject to significant measurement error. Appendix A.11 compares the median wage in my sample in the British Skill Survey and in the larger Annual Survey of Earnings and Hours (National Statistics (2018)). Moreover, it compares the city wage premium and its interaction with skill across both surveys.

[43]Negative hourly wage, hourly wage smaller than 1 and larger than 100.

Table 1.14: Interaction: city size, advanced computer use and university on individual wages in Great Britain

Dep.var. Ln of gross hourly wages$_i$	(1)	(2)	(3)	(4)	(5)
Indep. var.	OLS ASHE	OLS BSS	IV BSS	OLS BSS	IV BSS
Ln pop	.053***	.033***	.031**	.031***	.033***
	(.0107)	(.0089)	(.0120)	(.0085)	(.0103)
Uni	.44***	.30**	.25*	.57***	.53***
	(.034)	(.126)	(.130)	(.163)	(.176)
Uni × Ln pop	.016***	.015	.019*	-.005	-.002
	(.0026)	(.0099)	(.0099)	(.0129)	(.0139)
Advanced PC				.57***	.71***
				(.197)	(.228)
Advanced PC × Ln pop				-.020	-.031*
				(.0157)	(.0179)
Uni × Advanced PC				-1.003***	-1.050***
				(.313)	(.339)
Uni × Advanced PC × Ln pop				.062**	$\boxed{.066^{**}}$
				(.0245)	(.0268)
Observations	243,043	4,117	4,117	4,117	4117
Cragg-Donald F statistics			7780		3791

Std. Err. adjusted for 178 clusters (TTWA)
*** p<0.01, ** p<0.05, * p<0.1
Source: author's computation based on the 2006th wave of the British Sill Survey (BSS) as well as the 2006th wave of the Annual Survey of Hours and Earnings(ASHE).
Notes: The log of population is the log of population by travel to work area which represents an integrated labour market based on commuting flows. Advanced computer is a dummy variable taking value 1 if the worker has an advanced computer use at work, 0 otherwise. Instrument for log population in 1991: population in 1851 (census data).

The first two columns of Table 1.14 show the city wage premium as well as its interaction with being employed in an occupation requiring a university degree. Column (1) shows the city wage premium using data from the large Annual Survey of Hours and Earnings whose primary purpose is to document hours and wages.[44] I can thus compare it with column (2) which shows the same regression for the British Skill Survey. The main difference between the two surveys is that the interaction between university and population is insignificant in the BSS, whereas it is significant in the ASHE.[45] The magnitude of the coefficients is though remarkably similar with larger standard error for the smaller BSS. Next, column (3) instruments city size with historical population in the same regression as a robustness. Column (4) and (5) of Table 1.14 show that the coefficient of the triple interaction η_7 is positive and significant. An increase in city size of 10% is associated with an increase of .66% in the wage gap between workers in occupations requiring a university degree versus not, when both groups make an advanced computer use at work. Since I do not instrument advanced computer use at the individual level, column (5) still suffers from potential endogeneity.

Alternatively, Appendix A.12 shows that firms' labor demand for occupations inten-

[44]See Appendix A.11 for a description of the ASHE. This survey does not contain information on computer use at work.

[45]A significant interaction between city size and education on wages is in line with previous literature, see Combes and Gobillon (2015) for a review.

sive in advanced computer - that is with a large proportion of workers with an advanced computer use - relative to other occupations as measured by their share if the firms' wage bill increases with city size.

1.6 Conclusion

This paper formalizes the complementarity between computer capital and city size in a spatial equilibrium setting where workers are perfectly mobile across location. Because high-skilled workers disproportionately benefit from this complementarity, they choose to live and work in large agglomerations more often than other groups of workers. Consistent with the model I find that firms allocate more computerized capital as city size increases to workers of all skill types, not only to high-skilled workers. I find that a German worker with mean characteristics has a probability of 0.24 instead of 0.20 to have an advanced computer use at work in large urban centers relative to the rest of the country. Likewise, a British worker with mean characteristics has a probability of 0.19 instead of 0.16 to have an advanced computer use at work in large urban centers relative to the rest of the country. I also find that an increase in city size of 10% is associated with an increase of 0.62% in the wage gap between workers in occupation requiring a university degree and occupations not requiring it, when both groups make an advanced use of computers at work. In this context, the dispersion force related to recent technological change does not arise in the labor market, rather it acts through higher price on the rental market following the increasing attractiveness of large cities for high-skilled workers. If agglomeration or dispersion forces of technological change dominate might consequently crucially depend on housing policy and constraints.

Chapter 2

Computerization and urban centers: sorting of high-skilled workers and wage inequality

2.1 Introduction

This chapter seeks to quantify the impact of computerization on the sorting of high skilled workers in large cities. By computerization I mean a decline in the price of computerized equipment relative to other equipment and an increase in its relative productivity. Given the documented increasing sorting of high-skilled workers in large cities[1], I ask what would have been the change in the share of high-skilled workers in urban centers absent the decline in the relative price of computerized equipment. I also ask what would have been the wage inequality as a function of city size absent this decline.[2] As skill sorting tends to smooth wage differential across locations, a stronger change in the local skill distribution goes hand in hand with a lower local wage differential across skill types.

For this purpose, I extend the framework of Burstein et al. (2019) about workers' skill-equipment-occupation complementarities to include a spatial dimension with settlement types.[3] Workers choose their occupation, equipment type and location according to: (i) skill specific productivities or preferences, (ii) idiosyncratic productivities or preferences randomly distributed within skill groups, (iii) occupation and equipment prices as well as local rents. Workers' allocation across occupations, equipment and sites reflects production complementarities which are taken as given by the economic agents. This framework is flexible so that it can be taken to the data to perform numerical simulations. Whereas Chapter 1 models a possible source for the complementarity between computerized equipment and city size, the current chapter takes it as given. It seeks to quantify its impact on skill sorting, wherever its source.

[1]See e.g. Berry and Glaeser (2005) for the US.

[2]Baum-Snow and Pavan (2013) bring to light the emergence of a positive relationship between nominal wage inequality in the US since the 80th. Baum-Snow et al. (2018) and Lindley and Machin (2014) discuss some explanations behind it. See also Wheeler (2001) and Wheeler (2004) for a discussion about city population and inequality.

[3]Their assignment model with many groups of workers and occupations builds on Eaton and Kortum (2002), Lagakos and Waugh (2013) Lagakos and Wangh (2013) and Hsieh et al. (2019) and extents it to include many types of equipment.

Quantifying the impact of computerization on the share of high-skilled workers in large cities is difficult in the framework of Chapter 1. Indeed, its model is very stylized and although leading to sharp empirical predictions, lacks the flexibility to be parameterized empirically. On the contrary, the current model is flexible in that the complementarities are not theoretically stated but recovered empirically and that I construct wages and preferences for location with residual terms, which allows to match empirical data for wages and local population. Moreover, the structure of the model allows me to retrieve the relative price of computerized equipment, which is challenging to measure directly.

Next, I construct the model so that the change in population's type across settlements driven by changes in local unobserved skill is not falsely attributed to computerization. Indeed, I express the change in the proportion of high-skilled workers in large cities as a function of (i) the change in the relative price of computerized equipment common to all workers and (ii) measures of workers' specialization across equipment types in the first period available in the data. Change in local unobserved skill is captured by a residual term in the wage equation and is kept constant when performing the simulations. The method used is one way to test assignment models suggested by Acemoglu and Autor (2011) and addresses the drawbacks highlighted by DiNardo and Pischke (1997) when trying to capture the effect of computerization on wages. The key hypothesis for this is that the complementarity between equipment types, occupation and settlement types is time invariant.[4]

At equilibrium, a decrease in the relative price of computerized equipment affects local wage. First of all, it increases wages in settlements complementary with computerized equipment. If large cities, high skilled workers and computerized equipment are complement as highlighted by Chapter 1, local wages disproportionately increase in large cities for high skilled workers. A first key parameter governs the strength of the response of local wage to change in occupation and equipment prices and productivities: the dispersion of within groups idiosyncratic productivities for equipment and occupation. I identify this structural elasticity following Burstein et al. (2019) with the reduced form elasticity of changes in wages to changes in a weighted average of equipment and occupation prices and productivities. As occupation prices in wages are endogenous, I instrument the change in wage with a measure of change in equipment productivity.

Second, local wages are indirectly affected by a change in the price of occupation following the decrease in the price of computerized equipment. The output of occupations complementary with computerized equipment increases as workers relocate to these occupations. It increases their price and decreases the wage of workers employed in them. The elasticity of substitution across occupation governs the strength of this effect. I also identify this structural elasticity following Burstein et al. (2019) with the reduced form elasticity of occupation labor income to occupation price. Because of endogeneity of occupation prices, I use change in equipment price and productivity as an instrument.

[4]The model in Chapter 1 showed that population does affect the complementarity between city size and computer capital through agglomeration economies. The residual term capturing change in unobserved skill will thus also capture change in productivity due to change in city size. Because population is affected by computerization, my counterfactual estimate will possibly be biased. Including agglomeration economies though usually leads to multiple equilibria, which is problematic to perform counterfactuals.

Last, workers migrate within countries in response to changes in local wages.[5] The lower the dispersion in idiosyncratic preferences for location, the stronger the response, which is my last key elasticity. This feeds back in lower wage inequality as a function of city size through a decline in the price of occupation complementary with large cities. The reason is that its output increases when workers relocate there. I can quantify the extent of this feedback through the model. I identify this structural elasticity shaping the dispersion in idiosyncratic preferences for location using the reduced-form elasticity of the change in local population to the change in local wages. As occupation prices in wages are endogenous, I also instrument the change in wage with a measure of change in equipment productivity.

The second section describes the building blocks of the model, the third section discusses the equilibrium allocation and shows that the equilibrium exists and is unique. The fourth section expresses wages and the proportion of workers in a skill group in a city in changes over time and as a function of the change in the relative price of computerized equipment. section five illustrates how this system in difference can be parameterized with data from Germany. The last section discusses the impact of potential misspecifications of the model on quantifying effect of computerization on the sorting of high-skilled workers in large cities.

2.2 Elements of the model

This section presents the theoretical framework, where the notations stay as close as possible to Burstein et al. (2019).

2.2.1 Endowment

The economy consists of a continuum of workers with heterogeneous skill levels indexed by $z \in \mathcal{Z}_t$ divided into a finite number of groups $\lambda \in \mathcal{Z}_t(\lambda) \subseteq \mathcal{Z}_t$. The population of workers in group λ at time t in the economy is $l_t(\lambda)$, where the total population is $L_t = \sum_\lambda l_t(\lambda)$. There is a finite number of cities indexed by j that can host population. City's population is endogenous where $\xi_t(\lambda, j) \equiv \frac{l_t(\lambda,j)}{l_t(\lambda)}$ is the proportion of workers in a skill group λ choosing to work in city j.

2.2.2 Preferences

Equation (2.1) below describes the preferences of workers. They have a Cobb-Douglas utility function in housing $H_t(j, \lambda, z)$ and a fully tradable final good $C_t(j, \lambda, z)$. Utility is also a function of local amenities $Q_t(\lambda, j)$ which can differ across skill groups as they sometimes sort to different areas within cities and that amenities can be very local in nature.

[5]Since I consider an extended period, the change in the share of high-skilled workers in urban centers might also be driven by changes in education choices across city size as well as international migration non uniformly sorting across locations.

Moreover, there is also a within group heterogeneity in preferences for location $\psi_t(z, j)$. This encompasses idiosyncratic attachment to location such as family ties. Every worker receives a realization of the random variables $\psi_t(z, j)$ for every city j, independently from her skill group λ.

$$U_t(j, \lambda, z) = \frac{C_t(j, \lambda, z)^{1-\beta(\lambda)} H_t(j, \lambda, z)^{\beta(\lambda)}}{\beta(\lambda)^{\beta(\lambda)}(1-\beta(\lambda))^{(1-\beta(\lambda))}} Q_t(\lambda, j)^{\varrho(\lambda)} \psi_t(z, j) \tag{2.1}$$

Utility is subject to the budget constraint $\nu_t(z, \lambda, \kappa, \omega, j) = P_t C_t(j\lambda, z) + p_t(j) H_t(j, \lambda, z)$, where $\nu_t(z, \lambda, \kappa, \omega, j)$ is the wage of worker z in group λ and city j given the choice of her equipment κ and her occupation ω. P_t is the price of the final good, chosen as the numéraire and $p_t(j)$ is the rent in city j - its “price”. This price is paid to absentee landlords. $\beta(\lambda)$ is the share of income paid in order to live in city j. I allow it to vary with the skill group λ.[6] The preference for local amenities $\varrho(\lambda)$ varies with the skill groups λ as valuation for local amenities is correlated with education, for instance high-educated workers might value museums more than other groups or feel more legitimate visiting them.

The indirect utility of a worker z in group λ and city j is given by:

$$V_t(\lambda, z, \kappa, \omega, j) = \frac{\nu_t(z, \lambda, \kappa, \omega, j) Q_t(\lambda, j)^{\varrho(\lambda)} \psi_t(z, j)}{p_t(j)^{\beta(\lambda)} P_t^{1-\beta(\lambda)}} \tag{2.2}$$

Workers choose their location in order to maximize utility and choose their equipment at work κ and an occupation ω given the choice of the city such that wage is maximized.

2.2.3 Technology

Production in city j at time t by perfectly competitive firms is governed by:

$$Y_t(j) = \left(\sum_{\omega} \phi_t(\omega)^{\frac{1}{\rho}} Y_t(\omega, j)^{\frac{\rho-1}{\rho}} \right)^{\frac{\rho}{\rho-1}} \tag{2.3}$$

where $Y_t(\omega, j)$ is the fully tradable output produced in occupation ω in city j, where $\rho > 0$ is the elasticity of substitution across occupations and $\phi_t(\omega)$ is an occupation-specific demand shifter.

[6]It is a variation of housing preferences across skill groups. If the variation in income within skill groups across cities is small compared with the variation of income across skill group, it allows to consider the impact of distinctive shares of income allocated to housing on sorting. Preferences are though not strictly speaking non-homothetic. Even though I will set $\beta(\lambda) \equiv \beta \ \forall \lambda$ later in the analysis, I model preferences in this way to keep track of the bias introduced by my simplification.

Occupation output $Y_t(\omega, j)$ is produced by perfectly competitive units. A skill unit λ in occupation ω that employs k units of equipment κ and l efficiency units of labor produces:

$$k^{\alpha} \left[T_t(\lambda, \kappa, \omega, j) l\right]^{1-\alpha} \tag{2.4}$$

where $T_t(\lambda, \kappa, \omega, j)$ is the productivity of the efficiency units of labor and reflects the complementarity between the skill group λ, the equipment κ, the occupation ω and the city j.

In particular, a skill group λ has a comparative advantage over any other group λ' using equipment type κ if the productivity of this group is larger when it is equipped with κ compared to any other equipment κ' *relative* to a given skill group λ': $\frac{T_t(\lambda,\kappa,\omega,j)}{T_t(\lambda,\kappa',\omega,j)} > \frac{T_t(\lambda',\kappa,\omega,j)}{T_t(\lambda',\kappa',\omega,j)}$.

Alternatively, a city j has a comparative advantage over any other city j' using equipment type κ if the productivity in this city is larger when workers are equipped with κ compared to any other equipment κ' *relative* to a given city j': $\frac{T_t(\lambda,\kappa,\omega,j)}{T_t(\lambda,\kappa',\omega,j)} > \frac{T_t(\lambda,\kappa,\omega,j')}{T_t(\lambda,\kappa',\omega,j')}$. This is symmetric for other complementarities.

Moreover, every worker draws a realization of the random variables $\varepsilon_t(z, \kappa, \omega)$ for any occupation-equipment pairs (κ, ω), independently from her skill group λ. This represents her idiosyncratic productivity or preference for occupations and equipment. The distribution of $\varepsilon_t(z, \kappa, \omega)$ is independent from the distribution of idiosyncratic preferences for location $\psi_t(z, j)$.

Last, the resource constraint imposes that:

$$Y_t = C_t + \sum_{\kappa} p_t(\kappa) Y_t(\kappa) \tag{2.5}$$

where $Y_t \equiv \sum_j Y_t(j)$ and $Y_t(\kappa)$ (aggregate quantity of equipment type κ) is fully tradable. $Y_t(\kappa)$ fully depreciates each period. C_t describes aggregate consumption. $p_t(\kappa)$ is the price of equipment κ relative to the final good. Equipment are produced in a fully competitive environment with constant marginal cost.

2.3 Closing the model

2.3.1 Partial equilibrium

In this subsection, I take the price of occupations $p_t(\omega)$ and the price of cities $p_t(j)$ as given. I first present the timing of the decisions of workers:

I. Workers receive their idiosyncratic preferences for each city $\psi_t(z, j)$.

II. Workers choose their cities of dwelling according to their expected utility there.

III. Workers receive their idiosyncratic productivities $\varepsilon_t(z,\kappa,\omega)$ for each occupation-equipment pair and cannot relocate across cities at that point.

IV. Workers chose their occupation-equipment pair in order to maximize their wage given the choice of the city.

As I solve the model by backward induction, I first detail IV and then II. The timing repeats itself each period with a new generation of workers. Most of the calculations are left for Appendix B.

First of all, given the choice of the city, workers choose their occupation as well as their work equipment (κ,ω) by maximizing their wage $\nu(z,\lambda,\kappa,\omega,j) = v_t(\lambda,\kappa,\omega,j)\varepsilon(z,\kappa,\omega)$ in city j. I next show how to compute the term common for any worker in group λ, denoted $v_t(\lambda,\kappa,\omega,j)$ when choosing the occupation pair (κ,ω):

The revenue of an occupation production unit employing l and k units of κ is equal to $p_t(\omega)k^\alpha \left[T_t(\lambda,\kappa,\omega,j)l\right]^{1-\alpha}$. Next, the cost incurred by that occupation production unit is $p_t(\kappa)k + v_t(\lambda,\kappa,\omega,j)l$, where $v_t(\lambda,\kappa,\omega,j)$ is the wage per efficiency units of labor of type λ using equipment type κ in occupation ω and city j. Given these expression, the first order condition for the optimal quantity k of equipment κ in that occupation is:

$$k_t(.) = \left(\alpha\frac{p_t(\omega)}{p_t(\kappa)}\right)^{\frac{1}{1-\alpha}} T_t(\lambda,\kappa,\omega,j)l$$

Plugging this expression in the cost and revenue of the occupation production function and setting both equal to each others due to costless entry into occupation (zero profit condition) leads to:

$$v_t(\lambda,\kappa,\omega,j) = \bar{\alpha}p_t(\kappa)^{-\frac{\alpha}{1-\alpha}}p_t(\omega)^{\frac{1}{1-\alpha}}T_t(\lambda,\kappa,\omega,j). \tag{2.6}$$

The probability $\pi_t(\lambda,\kappa,\omega \mid j)$ that a randomly sampled worker $z \in \mathcal{Z}_t$ in group λ chooses κ and ω in city j depends on $v_t(\lambda,\kappa,\omega,j)$ as well as on her idiosyncratic productivity $\varepsilon_t(z,\kappa,\omega)$. I assume this random term to be Fréchet distributed with parameter $\theta > 1$, where a higher value of θ implies a lower within workers' λ group dispersion of efficiency units across occupations and equipment. In this case workers are relatively mobile across them. This distribution assumption leads to a tractable expression for $\pi_t(\lambda,\kappa,\omega \mid j)$:

$$\pi_t(\lambda,\kappa,\omega \mid j) = \frac{\left[T_t(\lambda,\kappa,\omega,j)p_t(\kappa)^{-\frac{\alpha}{1-\alpha}}p_t(\omega)^{\frac{1}{1-\alpha}}\right]^\theta}{\sum_{\kappa',\omega'}\left[T_t(\lambda,\kappa',\omega',j)p_t(\kappa')^{-\frac{\alpha}{1-\alpha}}p_t(\omega')^{\frac{1}{1-\alpha}}\right]^\theta} \tag{2.7}$$

Proof: Appendix B.1.

Furthermore, the average wage in group λ and city j is:

$$w_t(\lambda, j) = \gamma(\theta)\bar{\alpha}\left\{\sum_{\kappa',\omega'}\left[T_t(\lambda, \kappa', \omega', j)p_t(\kappa')^{-\frac{\alpha}{1-\alpha}}p_t(\omega')^{\frac{1}{1-\alpha}}\right]^{\theta}\right\}^{\frac{1}{\theta}} \tag{2.8}$$

where $\gamma(\theta) \equiv \Gamma(1 - \frac{1}{\theta})$ and $\Gamma(.)$ is the Gamma function and $\bar{\alpha} = (1-\alpha)\alpha^{\frac{\alpha}{(1-\alpha)}}$.

Proof Appendix B.2.

Second, the probability that a randomly sampled worker in group λ chooses city j is:

$$\xi_t(\lambda, j) = \frac{\left[\frac{w_t(\lambda,j)Q_t(\lambda,j)^{\varrho(\lambda)}}{p_t(j)^{\beta(\lambda)}}\right]^{\eta}}{\sum_{j'}\left[\frac{w_t(\lambda,j')Q_t(\lambda,j')^{\varrho(\lambda)}}{p_t(j')^{\beta(\lambda)}}\right]^{\eta}} \tag{2.9}$$

where η is the parameter of the distribution of the random idiosyncratic preferences for location $\psi_t(z, j)$, which is distributed Fréchet. A high value for η encompasses a low within group variation in preferences for location. As a result, local wages adjusted for the cost of living play a large role in the location choice. Workers are intrinsically relatively indifferent about locations.

Proof: Appendix B.3.

If workers could relocate they would have an incentive to. Indeed, the non-random part of their wage $v_t(\lambda, \kappa, \omega, j)$ in a city depends on the choice of occupation and equipment, which itself depends on the $\varepsilon_t(z, \kappa, \omega)$. If a worker receives an occupation-equipment draw which is complementary with a given city j, the probability that the indirect utility there is maximized increases as does her incentive to relocate to that city.

2.3.2 General equilibrium

The last two endogenous variables to recover are the price of the city $p_t(j)$ as well as the price of occupation $p_t(\omega)$.

City price

The price of the city is a function of the population in the city and its average wage there:

$$p_t(j) = \frac{\sum_{\lambda}\beta(\lambda)w_t(\lambda, j)}{\bar{H}} \tag{2.10}$$

Since the revenue from cities goes to absentee landlords, the model is partial in this respect.

Assuming for simplicity that $\beta(\lambda)$ is independent of λ, equation (2.10) can be rewritten as $p_t(j) = \beta \bar{w}_t(j)\frac{L_t(j)}{H}$, where $\bar{w}_t(j) = \sum_\lambda w_t(\lambda, j)\xi_t(\lambda, j)$ and $L_t(j) = \sum_\lambda L_t(\lambda)\xi_t(\lambda, j)$. Using equation 2.9 it becomes:

$$p_t(j) = p_t(j)^{-\beta\eta 2} \sum_\lambda \frac{w_t(\lambda, j)^{1+\eta} Q_t(\lambda, j)^{\varrho(\lambda)\eta}}{\sum_{j'} \left[\frac{w_t(\lambda,j')Q_t(\lambda,j')}{p_t(j')^\beta}\right]^\eta} \sum_\lambda \frac{L_t(\lambda) w_t(\lambda, j)^\eta Q_t(\lambda, j)^{\varrho(\lambda)\eta}}{\sum_{j'} \left[\frac{w_t(\lambda,j')Q_t(\lambda,j')}{p_t(j')^\beta}\right]^\eta}$$

$$\Leftrightarrow p_t(j) = \left\{ \sum_\lambda \frac{w_t(\lambda, j)^{1+\eta} Q_t(\lambda, j)^{\varrho(\lambda)\eta}}{\sum_{j'} \left[\frac{w_t(\lambda,j')Q_t(\lambda,j')}{p_t(j')^\beta}\right]^\eta} \sum_\lambda \frac{L_t(\lambda) w_t(\lambda, j)^\eta Q_t(\lambda, j)^{\varrho(\lambda)\eta}}{\sum_{j'} \left[\frac{w_t(\lambda,j')Q_t(\lambda,j')}{p_t(j')^\beta}\right]^\eta} \right\}^{\frac{1}{1+2\beta\eta}} \tag{2.11}$$

where $w_t(\lambda, j) = \gamma(\theta)\bar{\alpha} \left\{ \sum_{\kappa',\omega'} \left[T_t(\lambda, \kappa', \omega', j) p_t(\kappa')^{-\frac{\alpha}{1-\alpha}} p_t(\omega')^{\frac{1}{1-\alpha}} \right]^\theta \right\}^{\frac{1}{\theta}}$

Note that the right hand side of equation (2.11) depends on city j only trough the parameters $T_t(\lambda, \kappa', \omega', j)$ and $Q_t(\lambda, j)^{\varrho(\lambda)}$.

Occupation price

Next, I consider finding the zeros in the market for occupation ω as an excess demand system $Z_t(\omega, \boldsymbol{p_t(\omega)}) = 0$, where $\boldsymbol{p_t(\omega)} = (p_t(1), ..., p_t(n))$, where n is the number of occupations. This will be convenient to prove the existence of the equilibrium in the next section.

On the demand side for the market for occupation, the first order condition for the final firm implies that the total expenditure of the firm in a given city on the occupation is equal to:

$$y_t(\omega, j)p_t(\omega) = Y_t(j)p_t(\omega)^{1-\rho}\phi_t(\omega)P_t^\rho \tag{2.12}$$

with

$$y_t(\omega, j) = Y_t(j)\phi_t(\omega)\left[\frac{P_t}{p_t(\omega)}\right]^\rho \tag{2.13}$$

where P_t is equal to the marginal cost of the representative firm in each city due to perfect competition and is equal to $P_t = \left[\sum_\omega \phi_t(\omega)p_t(\omega)^{1-\rho}\right]^{\frac{1}{1-\rho}}$.

Moreover, $P_tY_t \equiv P_t\sum_{j,\omega} y_t(\omega, j) = E_t$ in equilibrium, because final firms make no profit, where E_t is the total labor income in the economy. Because of the Cobb-Douglas structure of the production of occupation, labor income $= (1-\alpha)E_t$. Thus, $E_t = \dfrac{1}{1-\alpha}\sum_{j,\lambda} w_t(j, \lambda)L_t(\lambda)\xi_t(\lambda, j)$.

Summing equation (2.12) over all cities gives the total expenditure in occupation ω in the economy: $y_t(\omega)p_t(\omega) = E_t p_t(\omega)^{1-\rho} P_t^{\rho-1} \sum_j \phi_t(\omega)$.

On the supply side for the market for occupation, the revenue earned by the skill units producing ω in all cities, where $p_t(\omega)$ is the same across cities due to perfect tradability: $y_t(\omega)p_t(\omega) = \dfrac{1}{1-\alpha}\zeta_t(\omega)$,

where $\zeta_t(\omega) = \sum_{j,\lambda,\kappa} w_t(j,\lambda)L_t(\lambda)\xi_t(\lambda,j)\pi_t(\lambda,\kappa,\omega,j)$ is the labor income in occupation ω.

Dividing the expenditure and revenue in an occupation by the occupation price $p_t(\omega)$ describes the market for occupation as an excess demand system $Z_t(\omega, p_t(\omega)) = 0$:

$$Z_t(\omega, p_t(\omega)) = \frac{1}{p_t(\omega)}\left[E_t p_t(\omega)^{1-\rho} P_t^{\rho-1}\phi_t(\omega) - \frac{1}{1-\alpha}\zeta_t(\omega)\right] \ \forall\, \omega \tag{2.14}$$

Definition of the equilibrium

An equilibrium is an occupation price vector $\boldsymbol{p_t(\omega)}$ and a city price vector $\boldsymbol{p_t(j)}$ such that $Z_t(\omega, p(\omega)) = 0$, $\forall\ \omega = 1, ..., n$ and satisfying 2.11 $\forall\ j = 1, ..., m$ in all time periods.[7] Given both of these vectors, all other equilibrium prices and quantities can be computed as described in Section 2.3.1. The aggregate quantity of equipment κ, denoted $y_t(\kappa)$, is given by equating the total income $\vartheta_t(\kappa)$ earned by production factors associated to equipment type κ in all cities from the perspective of labor and capital: $\vartheta_t(\kappa) = \dfrac{1}{\alpha}$capital income $= \dfrac{1}{1-\alpha}$labor income:

$$y_t(\kappa) = \frac{1}{p_t(\kappa)}\frac{\alpha}{1-\alpha}\sum_{j,\lambda,\omega} w_t(j,\lambda)L_t(\lambda)\xi_t(\lambda,j)\pi_t(\lambda,\kappa,\omega,j) \quad \forall\, \kappa = 1, ..., l \tag{2.15}$$

Last, aggregate consumption is given by equation (2.5).

2.3.3 Existence and uniqueness of the equilibrium

In this section, I prove the existence and uniqueness of the equilibrium staying as close as possible to the logic of Section 4 from Alvarez and Lucas (2007).

Assumption 2 *I.* $0 < \beta < 1$, $\theta, \eta > 1$,

[7] Since capital fully depreciates each period, the problem can be solved independently for each period.

II. for some numbers $\underline{Q}$, $\underline{T}$ *and* $\underline{\phi}$[8], $0 < \underline{Q} \leq Q_t(\lambda, j)^{\varrho(\lambda)} \leq 1$,

III. and $0 < \underline{T} \leq T_t(\lambda, \kappa, \omega, j) \leq 1$,

IV. and $0 < \underline{\phi} \leq \phi_t(\omega) \leq 1$.

Lemma 4 *Under assumptions 2, for any* $\boldsymbol{p_t(\omega)} \in \mathbf{R}^n_{++}$ *there is a unique vector* $\boldsymbol{p_t(j)}$ *that satisfies equation* (2.11)*. For each city* j*, the function* $p_t(j, \boldsymbol{p_t(\omega)})$ *is:*

I. continuously differentiable in $\boldsymbol{p_t(\omega)}$,

II. homogeneous of degree one in prices $\boldsymbol{p_t(\omega)}$ *and* $\boldsymbol{p_t(\kappa)}$,

III. strictly increasing in $\boldsymbol{p_t(\omega)}$,

IV. strictly increasing in the parameters $T_t(\lambda, \kappa, \omega, j)\ \forall\ \lambda, \kappa, \omega$ *and* $Q_t(\lambda, j)\ \forall\ \lambda$ *and strictly decreasing in* $T_t(\lambda, \kappa, \omega, j'')\ \forall\ \lambda, \kappa, \omega$ *and* $Q_t(\lambda, j'')\ \forall\ \lambda,\ j \neq j''$,

V. satisfies the bounds $\underline{p}(j, \boldsymbol{p_t(\omega)}) \leq p_t(j, \boldsymbol{p_t(\omega)}) \leq \overline{p}(j, \boldsymbol{p_t(\omega)})$ *for all* $p_t(j, \boldsymbol{p_t(\omega)}) \in \mathbf{R}^n_{++}$*, where*

$$\underline{p}(j, \boldsymbol{p_t(\omega)}) = \Lambda^2 L_t \left[\frac{(m-1)\overline{w}(\boldsymbol{p_t(\omega)})^{\frac{\eta(1-\beta)}{1+2\beta\eta}}}{\underline{w}(\boldsymbol{p_t(\omega)})^{\frac{1+2\eta}{2(1+2\beta\eta)}} \underline{Q}^{\frac{\eta(1+\beta\eta)}{1+2\beta\eta}}} + \underline{w}(\boldsymbol{p_t(\omega)})^{-1/2} \right]^{-2}$$

and

$$\overline{p}(j, \boldsymbol{p_t(\omega)}) = \Lambda^2 L_t \left[\frac{(m-1)\underline{w}(\boldsymbol{p_t(\omega)})^{\frac{\eta(1-\beta)}{1+2\beta\eta}} \underline{Q}^{\frac{\eta(1+\beta\eta)}{1+2\beta\eta}}}{\overline{w}(\boldsymbol{p_t(\omega)})^{\frac{1+2\eta}{2(1+2\beta\eta)}}} + \overline{w}(\boldsymbol{p_t(\omega)})^{-1/2} \right]^{-2}$$

$$\underline{w}(\boldsymbol{p_t(\omega)}) = \gamma(\theta)\bar{\alpha} \left\{ \textstyle\sum_{\kappa', \omega'} \left[\underline{T} p_t(\kappa')^{-\frac{\alpha}{1-\alpha}} p_t(\omega')^{\frac{1}{1-\alpha}} \right]^{\theta} \right\}^{\frac{1}{\theta}}$$

and $\overline{w}(\boldsymbol{p_t(\omega)}) = \gamma(\theta)\bar{\alpha} \left\{ \sum_{\kappa', \omega'} \left[p_t(\kappa')^{-\frac{\alpha}{1-\alpha}} p_t(\omega')^{\frac{1}{1-\alpha}} \right]^{\theta} \right\}^{\frac{1}{\theta}}$.

Proof: Appendix B.4.

Note that the hypothesis that $\beta(\lambda) \equiv \beta\ \forall \beta$ is crucial to prove homogeneity in prices of the function $p_t(j, \boldsymbol{p_t(\omega)})$ (part II, Lemma 4).

[8]Normalization of units.

Proposition 3 *Under Assumptions 2 and with $P_t = 1$, there is a $\boldsymbol{p_t(\omega)} \in \mathbf{R}^n_{++}$ such that $Z_t(\omega, \boldsymbol{p_t(\omega)}) = 0$*

I. $Z_t(\omega, \boldsymbol{p_t(\omega)})$ *is continuous,*

II. $Z_t(\omega, \boldsymbol{p_t(\omega)})$ *is homogeneous of degree zero in prices* $\boldsymbol{p_t(\omega)}$, $\boldsymbol{p_t(\kappa)}$ *and* P_t,

III. $p_t(\omega) Z_t(\omega, p_t(\omega)) = 0 \; \forall \; p_t(\omega)$ *and* $\omega \in \mathbf{R}^n_{++}$ *(Walras's Law),*

IV. *there is an* $s > 0$ *such that* $Z_t(\omega, p_t(\omega)) > -s$ *for every occupation* $\omega = 1, ..., n$ *and all* $p_t(\omega) < 1$, *which is a bound on* $p_t(\omega)$ *and where* $s = \left(\frac{\alpha}{p_t(\kappa)}\right)^{\frac{\alpha}{1-\alpha}} L_t$,

V. *if* $\boldsymbol{p_t(\omega)}^b \to \boldsymbol{p_t(\omega)}^0$, *where* $\boldsymbol{p_t(\omega)}^0 \neq 0$ *and* $p_t(\omega) = 0$ *for some* ω, *then* $\max_k \left\{ Z_t(k, \boldsymbol{p_t(\omega)}^b) \right\} \to \infty$.

The result then follows from Proposition 17.C.1 p. 585 in Mas-Colell et al. (1995).

Proof: Appendix B.5.

Proposition 4 *Under Assumptions 2, there is exactly one solution to $Z_t(\omega, \boldsymbol{p_t(\omega)}) = 0$ that satisfies $P_t = 1$ (numéraire), because Z has the gross substitute property:*

$$\frac{\partial Z_t(\omega, \boldsymbol{p_t(\omega)})}{\partial p_t(\omega'')} > 0 \; \forall \omega, \; \omega \neq \omega'', \; \forall \; p_t(\omega) \in \mathbf{R}^n_{++} \tag{2.16}$$

Proof: Appendix B.6.

The result then follows from Proposition 17.F.3 p. 613 in Mas-Colell et al. (1995).

2.4 Decomposing changes over time in local wages and population

In this section, I define changes in any variable x between any two period t_0 and t_1 by ($\hat{x} \equiv \frac{x_{t_1}}{x_{t_0}}$), following the notation of Burstein et al. (2019). I compute changes over time as I am interested in the change in local skill compositions and local wage inequality following computerization.

First of all, I compare the change in city population (relative to other cities) across skill groups λ:

$$\frac{\hat{l}(\lambda, j)}{\hat{l}(\lambda, j_1)} \frac{\hat{l}(\lambda_1, j_1)}{\hat{l}(\lambda_1, j)} \tag{2.17}$$

where $\hat{l}(\lambda, j) \equiv \hat{\xi}(\lambda, j)\hat{l}(\lambda)$ is the change over time of the population in skill group λ in city j as a function of the change in computer price.

To illustrate, let be only two skill groups λ_1 and λ_2, where the first group is workers with a university degree and the second workers without a university degree. Let also be only two settlement types j_1 and j_2, where the first is urban centers and the second the rest of the country. Equation (2.17) then describes the change in the share of workers with a university degree in urban centers relative to the rest of the country.

Second, I compute the local change in wage over time of a worker in group λ relative to a reference group as a function of the change in computer price:

$$\frac{\hat{w}(\lambda, j)}{\hat{w}(\lambda_1, j)} \tag{2.18}$$

I then compute a measure of inequality based on equation (2.18).

Last, I can quantify the smoothing effect of internal migration on wage inequality as a function of city size by computing wage inequality had workers not relocate following the decline in the relative price of computerized equipment.

2.4.1 Change in city population

Rewriting equation (2.9) in change over time between period 0 and 1 ($\hat{x}(j) \equiv \frac{x_{t_1(j)}}{x_{t_0}(j)}$) as a function of a reference group λ' and as a function of a reference city j' gives:

$$\frac{\hat{l}(\lambda, j)}{\hat{l}(\lambda, j')}\frac{\hat{l}(\lambda', j')}{\hat{l}(\lambda', j)} = \left[\frac{\hat{w}(\lambda', j')}{\hat{w}(\lambda, j')}\right]^{\eta}\left[\frac{\hat{w}(\lambda, j)}{\hat{w}(\lambda', j)}\right]^{\eta}\left[\frac{\hat{Q}(\lambda', j')^{\varrho(\lambda')}}{\hat{Q}(\lambda', j)^{\varrho(\lambda')}}\frac{\hat{Q}(\lambda, j)^{\varrho(\lambda)}}{\hat{Q}(\lambda, j')^{\varrho(\lambda)}}\right]^{\eta} \tag{2.19}$$

where $\frac{\hat{w}(\lambda,j)}{\hat{w}(\lambda',j)}$ is computed in the next subsection, η is to be estimated from the data and $\left[\frac{\hat{Q}(\lambda',j')^{\varrho(\lambda')}}{\hat{Q}(\lambda',j)^{\varrho(\lambda')}}\frac{\hat{Q}(\lambda,j)^{\varrho(\lambda)}}{\hat{Q}(\lambda,j')^{\varrho(\lambda)}}\right]^{\eta}$ is a residual to match observed changes in location and wage so that it will capture change in quality of life through local amenities as well as model misspecification or anything relevant that the model does not account for.

2.4.2 Change in local wages

Next, I compute the change in wage as a function of the change in computer prices.

Consider equation (2.8) and let

$$T_t(\lambda, \kappa, \omega, j) \equiv T_t(\lambda, j)T_t(\kappa)T_t(\omega)T(\lambda, \kappa, \omega, j) \tag{2.20}$$

In this specification, the comparative advantage between skill group, equipment, occupation and city $T(\lambda, \kappa, \omega, j)$ is time invariant and enters multiplicatively with the productivity of the skill group λ in city j, denoted $T_t(\lambda, j)$, the productivity of equipment κ, denoted $T_t(\kappa)$ and the productivity of occupation ω, denoted $T_t(\omega)$.

The average wage of a worker z in group $\lambda \in \mathcal{Z}_t(\lambda) \subseteq \mathcal{Z}_t$ in city j is given by:

$$\hat{w}(\lambda, j) = \hat{T}(\lambda, j) \left\{ \sum_{\kappa', \omega'} \left[\hat{T}(\omega')\hat{p}(\omega')^{\frac{1}{1-\alpha}} \hat{T}(\kappa')\hat{p}(\kappa')^{-\frac{\alpha}{1-\alpha}} \right]^{\theta} \pi_0(\lambda, \kappa', \omega', j) \right\}^{\frac{1}{\theta}} \tag{2.21}$$

Proof: see Appendix B.7

Equation (2.22) provides the change in factor allocation, rearranging equation (2.7), with $\hat{q}(\omega) \equiv \hat{T}(\omega)\hat{p}(\omega)^{\frac{1}{1-\alpha}}$, $\hat{q}(\kappa) \equiv \hat{T}(\kappa)\hat{p}(\kappa)^{-\frac{\alpha}{1-\alpha}}$.

$$\hat{\pi}(\lambda, \kappa, \omega, j) = \frac{[\hat{q}(\omega)\hat{q}(\kappa)]^{\theta}}{\sum_{\kappa', \omega'} [\hat{q}(\omega')\hat{q}(\kappa')]^{\theta} \pi_0(\lambda, \kappa', \omega', j)} \tag{2.22}$$

The change in wage of a group of worker λ relative to a group λ' in city j is given by:

$$\begin{aligned}
\frac{\hat{w}(\lambda, j)}{\hat{w}(\lambda', j)} &= \frac{\hat{T}(\lambda, j)\hat{q}(\omega_1)\hat{q}(\kappa_1) \left\{ \sum_{\kappa', \omega'} [\hat{q}(\omega')\hat{q}(\kappa')]^{\theta} \pi_0(\lambda, \kappa', \omega', j) \right\}^{\frac{1}{\theta}}}{\hat{T}(\lambda', j)\hat{q}(\omega_1)\hat{q}(\kappa_1) \left\{ \sum_{\kappa', \omega'} [\hat{q}(\omega')\hat{q}(\kappa')]^{\theta} \pi_0(\lambda', \kappa', \omega', j) \right\}^{\frac{1}{\theta}}} \\
&= \frac{\hat{T}(\lambda, j) \left\{ \sum_{\kappa', \omega'} \frac{\hat{q}(\omega')^{\theta}}{\hat{q}(\omega_1)^{\theta}} \frac{\hat{q}(\kappa')^{\theta}}{\hat{q}(\kappa_1)^{\theta}} \pi_0(\lambda, \kappa', \omega', j) \right\}^{\frac{1}{\theta}}}{\hat{T}(\lambda', j) \left\{ \sum_{\kappa', \omega'} \frac{\hat{q}(\omega')^{\theta}}{\hat{q}(\omega_1)^{\theta}} \frac{\hat{q}(\kappa')^{\theta}}{\hat{q}(\kappa_1)^{\theta}} \pi_0(\lambda', \kappa', \omega', j) \right\}^{\frac{1}{\theta}}}
\end{aligned} \tag{2.23}$$

where θ is to be estimated from the data, $\frac{\hat{T}(\lambda, j)}{\hat{T}(\lambda', j)}$ is a residual to match the observed change in local wages, so that it will capture change in labor-city productivity as well as model misspecification or anything relevant that the model does not account for, and $\frac{\hat{q}(\kappa')^{\theta}}{\hat{q}(\kappa_1)^{\theta}}$, $\frac{\hat{q}(\omega')^{\theta}}{\hat{q}(\omega_1)^{\theta}}$ are retrieved from equation (2.22):

$$\frac{\hat{\pi}(\lambda, \kappa, \omega, j)}{\hat{\pi}(\lambda, \kappa_1, \omega, j)} = \frac{\hat{q}(\kappa)^{\theta}}{\hat{q}(\kappa_1)^{\theta}} \tag{2.24}$$

$$\frac{\hat{\pi}(\lambda, \kappa, \omega, j)}{\hat{\pi}(\lambda, \kappa, \omega_1, j)} = \frac{\hat{q}(\omega)^{\theta}}{\hat{q}(\omega_1)^{\theta}} \tag{2.25}$$

Population $l(\lambda, j)$, wages $w(\lambda, j)$ and factor allocation $\pi(\lambda, \kappa, \omega, j)$ are observed in the data.

Last, I link the change in the endogenous price of occupation to the change in the relative equipment prices. Indeed, equation (2.23) depends on the later directly and indirectly through the price of occupation. For this purpose, I rearrange equation (2.14) in time difference (equation (2.26) below), substituting $\hat{p}(\omega)^{1-\rho} = \hat{q}(\omega)^{(1-\alpha)(1-\rho)}\hat{T}(\omega)^{(1-\alpha)(\rho-1)}$. All variables are observed in my datasets, except the occupation specific productivity change $\hat{a}(\omega) \equiv \hat{T}(\omega)^{(1-\alpha)(\rho-1)}\hat{\phi}(\omega)$, which I recover with that equation. Also, I need a value for the parameters α, and ρ. I can then construct the counterfactual price of occupation ω relative to a reference occupation ω_1 absent the change in the relative price of computerized equipment, where $\hat{E}$ cancels out.

$$\hat{E}\hat{q}(\omega)^{(1-\alpha)(1-\rho)}\hat{a}(\omega) = \hat{\zeta}(\omega) \tag{2.26}$$

where $\hat{\zeta}(\omega) = \dfrac{1}{\zeta_0(\omega)} \sum_{j,\lambda,\kappa} \hat{w}(j, \lambda)\hat{L}(\lambda)\hat{\xi}(\lambda, j)\hat{\pi}(\lambda, \kappa, \omega, j)w_0(j, \lambda)L_0(\lambda)\xi_0(\lambda, j)\pi_0(\lambda, \kappa, \omega, j)$.

2.4.3 Intuition

What happens to wages when the relative price of computerized equipment falls? First of all, they increase directly for skill groups λ in occupations and in city that has a comparative advantage with this equipment. That is where the initial allocation $\pi_0(\lambda, \kappa, \omega, j)$ is large compared to other labor groups, occupations and cities.

Second, wages are indirectly affected by the change in the price of occupations. Indeed, the price of occupations with a comparative advantage with computerized equipment diminishes as workers relocate to these occupations. This decreases the relative wage of every worker employed in this occupation relative to others, in all cities and for all work equipment. A low dispersion of within groups idiosyncratic productivities for equipment and occupations θ reinforce this effect. Because workers are equally productive across equipment and occupations, they switch easily. On the contrary, if θ is high, workers have strong preferences for specific equipment and occupations and are reluctant to change. Take as an example the introduction of computerized equipment in clerical occupation. Workers might have been very reluctant to change equipment. In this case, the wages of the few workers who accepted to change is larger than in the case where the vast majority of workers adopt computers.

Finally, wages are also indirectly impacted by the relocation of workers across settlement types. As workers relocate to cities complement with computerized equipment, they produce more output in their new settlements as they would elsewhere. The price of their occupation thus decreases as a feedback. This decreases their wage as well as the wage of every worker employed in this occupation relative to others, in every city and for every work equipment they use. The strength of this feedback depends on the dispersion of idiosyncratic preferences for location η. If η is low workers are more indifferent across location compared to a situation where η is high. Maybe computerized equipment made clerical occupation more productive in large cities. If workers are geographically mobile, they move in numbers to larger cities to enjoy the higher wage. As a result, the production of the fully tradable clerical output increases overall and its price decreases. This translates into a lower wage in the occupation. The strength of this feeback is small if workers are not mobile.

2.5 Parameterization

In order to take the model to the data, I need information on my observables which are: population by skill group and settlement type $l(\lambda, j)$, average wages by skill group and settlement type $w(\lambda, j)$, labor income by occupation $\zeta(\omega)$ and factor allocation for each skill group, occupation, equipment and settlement type $\pi(\lambda, \kappa, \omega, j)$.

I also need measures for the changes in relative equipment and occupation prices $\frac{\hat{q}(\kappa)^\theta}{\hat{q}(\kappa_1)^\theta}$ and $\frac{\hat{q}(\omega)^\theta}{\hat{q}(\omega_1)^\theta}$.

Last, I need an estimate of the dispersion of idiosyncratic preferences for location η as well as the elasticity of substitution between occupations ρ and the dispersion of within groups idiosyncratic productivities for equipment and occupations θ. I calibrate the share of income allocated to equipment α with estimates from the literature.[9]

2.5.1 Data

I use the six waves of the German working population survey conducted by the Federal institute for vocational education and training (BIBB) and the Federal Institute for Occupational Safety and Health (BAuA) (1979, 1986, 1992, 1999, 2006, 2012).[10] It gives detailed information on the type of equipment used at work and has between $20'000$ and $35'000$ observations depending on the wave. I combine it with wage data from the larger Sample of Integrated Labour Market Biographies (SIAB) available from 1975 onward, where I also take occupation shares and population.[11] The dataset is a 2% of administrative social records in Germany. As such it excludes workers not covered by the social security system. It is also right-censored at the highest level of earnings.[12] I link the two datasets by relating group averages based on education, occupation and type of settlement.

I divide the population in four education groups, where education is defined quite consistently over the 1979-2012 period: primary education, secondary 1 and 2, secondary 3 without a university degree and secondary 3 with a university degree (or tertiary education). Moreover, I will also consider sex and two age groups. I consider about ten

[9]E.g. Karabarbounis and Neiman (2013), Figure 2. Their estimate of α for Germany over the period ranges from .32 to .40

[10]I will either focus on West Germany or drop the first two waves.

[11]There were three (micro) census in West Germany around the period: 1970, 1987 and 2011. I will cross-check the information in the two samples depending on the variables available.

[12]https://fdz.iab.de/en/FDZ_Individual_Data/integrated_labour_market_biographies.aspx, see Dustmann et al. (2009) for a full data description as well as advantages and disadvantages of these data compared to other wage data in Germany.

occupations, where the classification of occupation has changed considerably across the period, with the introduction of the unified ISCO classification in 1988.[13] Moreover, I group equipment types based on a question indicating which work equipment is the main one. Indeed, in the model workers only choose one equipment type, which can be described as the main work equipment. The classification of equipment also changes over time, though computerized equipment are clearly identifiable. I create a dummy indicating if the main work equipment is a computerized equipment. Last, I use the German definition of settlements, which divides the German territory into three categories: urban areas, urbanizing regions and rural areas.[14] Even though their definitions has slightly changed over time, they remain comparable.

2.5.2 Change in computer and occupation prices

The change in computer price relative to other equipment is my measure of computerization. As equation (2.24) shows, it can be retrieved with the change in the probability that workers have computers denoted κ_2 as main work equipment relative to other equipment grouped as κ_1, in a given skill group, occupation and settlement type.

$$\frac{\hat{q}(\kappa_2)^\theta}{\hat{q}(\kappa_1)^\theta} = \frac{\hat{\pi}(\lambda, \kappa_2, \omega, j)}{\hat{\pi}(\lambda, \kappa_1, \omega, j)} = \frac{\#\text{ workers with computer as main equipment}}{\#\text{ workers with other main equipment}} \tag{2.27}$$

In practice, I compute the relative price of computerized equipment through a geometric average of the share of workers with computer as main work equipment in each skill group λ, occupation ω and settlement types j following the baseline procedure of Burstein et al. (2019) (where I additionally condition on settlement types, which they do not have):

$$\frac{\hat{q}(\kappa_2)^\theta}{\hat{q}(\kappa_1)^\theta} = \exp\left(\frac{1}{N(\kappa_1, \kappa_2)} \sum_{\lambda,\omega,j} log \frac{\hat{\pi}(\lambda, \kappa_2, \omega, j)}{\hat{\pi}(\lambda, \kappa_1, \omega, j)}\right) \tag{2.28}$$

In the absence of zero in the change in factor allocation $\hat{\pi}(\lambda, \kappa, \omega, j)$, $N(\kappa_1, \kappa_2)$ is the number of labor groups multiplied by the number of occupations and the number of settlement types.

One problem with equation (2.28) is that my skill, occupation and settlement groups are unlikely to encompass all relevant dimensions of complementarity with equipment. In particular, if workers increasingly sort to large cities for reason unrelated to computerization within groups according to ability, the change in relative factor allocation will not capture change in relative equipment price. Nevertheless, averaging for all groups in equation (2.28) mitigates the concern. Indeed, in large cities, relative factor allocation

[13] If I can have more occupations, I will make use of it, but given the change in classification, a fine time-consistent classification is not feasible.

[14] In German: Städtische Regionen, Regionen mit Verstädterungsansätzen, Ländliche Regionen. The two main criteria for those areas are population density and the presence of a large city center.

will overestimate the true change in relative equipment price for high-skilled groups. The reverse is true in rural areas. I will then check the soundness of the change in the local wage residual term $\frac{\hat{T}(\lambda,j)}{\hat{T}(\lambda,j_1)}$. I expect it to be increasing in the skill group in large cities relative to smaller ones.

I carry out a similar procedure to retrieve the change in occupation price (based on equation (2.25)), I first compute the price of each occupation ω relative to a given reference occupation ω_0:

$$\frac{\hat{q}(\omega)^\theta}{\hat{q}(\omega_0)^\theta} = \exp\left(\frac{1}{N(\omega,\omega_0)}\sum_{\lambda,\kappa,j} log\frac{\hat{\pi}(\lambda,\kappa,\omega,j)}{\hat{\pi}(\lambda,\kappa,\omega_0,j)}\right) \tag{2.29}$$

In the absence of zero in the change in factor allocation $\hat{\pi}(\lambda,\kappa_2,\omega,j)$, $N(\omega,\omega_0)$ is the number of labor groups multiplied by the number of equipment types and the number of settlement types.

In the presence of zeros in the change in factor allocation $\hat{\pi}(\lambda,\kappa_2,\omega,j)$, $N(\omega,\omega_0)$ varies with the choice of the reference occupations ω,ω_0. Taking the following geometric average helps reduce the dependence of the relative occupation price to the choice of a reference occupation ω_1:

$$\frac{\hat{q}(\omega)^\theta}{\hat{q}(\omega_1)^\theta} = \exp\left(\frac{1}{N(\omega_0)}\sum_{\omega_0}\left(\log\frac{\hat{q}(\omega)^\theta}{\hat{q}(\omega_0)^\theta} - \log\frac{\hat{q}(\omega_1)^\theta}{\hat{q}(\omega_0)^\theta}\right)\right) \tag{2.30}$$

where $N(\omega_0)$ is the number of occupations. This expression reduces to equation (2.29) in the case where no change of factor allocation is zero.

2.5.3 Estimating the dispersion of idiosyncratic preferences for location

I now expose the procedure to estimate the dispersion of idiosyncratic preferences for location η. For this purpose, I use the reduced-form elasticity of the change in relative local labor supply with respect to the change in local wages.

Rearranging equation (2.19) and taking log gives:

$$\ln\tilde{l}(\lambda,\lambda_1,j,j_1,t) = \eta\ln\tilde{w}(\lambda,\lambda_1,j,j_1,t) + \underbrace{\ln\tilde{Q}(\lambda,\lambda_1,j,j_1,t)^\eta}_{\text{unobserved}} \tag{2.31}$$

where $\tilde{l}(\lambda,\lambda_1,j,j_1,t) \equiv \frac{\frac{\hat{l}(\lambda,j,t)}{\hat{l}(\lambda,j_1,t)}}{\frac{\hat{l}(\lambda_1,j,t)}{\hat{l}(\lambda_1,j_1,t)}}$ is the relative change in the population of skill group λ in settlement type j relative to the reference group λ_1 and settlement type j_1, $\tilde{w}(\lambda,\lambda_1,j,j_1,t) \equiv \frac{\frac{\hat{w}(\lambda,j)}{\hat{w}(\lambda,j_1)}}{\frac{\hat{w}(\lambda_1,j)}{\hat{w}(\lambda_1,j_1)}}$ the relative change in the average wage of skill group λ in settlement type j relative to the reference group λ_1 and settlement type j_1 and $\tilde{Q}(\lambda,\lambda_1,j,j_1,t) \equiv \frac{\left[\frac{\hat{Q}(\lambda,j,t)}{\hat{Q}(\lambda,j_1,t)}\right]^{\varrho(\lambda)}}{\left[\frac{\hat{Q}(\lambda_1,j,t)}{\hat{Q}(\lambda_1,j_1,t)}\right]^{\varrho(\lambda_1)}}$ the relative change in the perceived quality of life of skill group λ in settlement type j relative to the reference group λ_1 and settlement type j_1.

I estimate η based on equation (2.31), where the change in relative quality of life $\tilde{Q}(\lambda,\lambda_1,j,j_1,t)^{\eta}$ is part of the error term. Indeed, I need a value for η to recover it. The dimensionality of equation (2.31) is the number of skill groups λ minus 1, the number of settlement types j minus 1 and the number of time periods t minus 1. Appendix B.8 discusses the choice of the reference groups.

Equation (2.31) suffers from reverse causality as relative wages across locations are endogenous through general equilibrium effects. Indeed, occupation price $\frac{\hat{q}(\omega')^{\theta}}{\hat{q}(\omega_1)^{\theta}}$ in equation (2.14) depends on the share of workers in group λ in each city $\xi(\lambda,j) \equiv \frac{l(\lambda,j)}{l(\lambda)}$. If the share of workers employed in an occupation in which the city has a comparative advantage increases, the price of that occupation decreases. This in turns leads wages to decrease in this city relative to other cities.

Moreover, the unobserved change in quality of life $\hat{Q}(\lambda,j_1)^{\eta}$ is correlated with local wages through population in the occupation price (equations (2.26), 2.23 and 2.19). It leads to an omitted variable bias. Also, other mechanisms neglected by the model can be a source of bias. For example, the perceived quality of life can affect motivation for instance, which affects productivity. As I take my model to the data, these effects would be captured by the residual term $\hat{T}(\lambda,j)$ in the wage equation (2.23). Moreover, taxes levied locally are positively linked to local wages, which allows investment in infrastructures which can improve quality of life. The perception of the utility of these infrastructures can be different for different skill groups. Also, local wages can directly impact quality of life if residents do not want to have poor workers around them or on the contrary if they enjoy it by feeling richer compared to them. The expected direction of the bias is thus overall non-determined.

Education fixed-effects might help to reduce these concerns if the dimensionality of the problem allows for it. It nevertheless will not completely solve it. I thus also construct the following Barthik-like instrument for $\tilde{w}(\lambda,\lambda_1,j,j_1,t)$:

$$\tilde{\mathcal{X}}_{\eta}(\lambda,\lambda_1,j,j_1,t) = \frac{\frac{\mathcal{X}_{\eta}(\lambda,j,t)}{\mathcal{X}_{\eta}(\lambda,j_1,t)}}{\frac{\mathcal{X}_{\eta}(\lambda_1,j,t)}{\mathcal{X}_{\eta}(\lambda_1,j_1,t)}} \tag{2.32}$$

where $\mathcal{X}_\eta(\lambda, j, t) = \sum_\kappa \frac{\hat{q}(\kappa,t)^\theta}{\hat{q}(\kappa_1,t)^\theta} \sum_\omega \pi_{1979}(\lambda, \kappa, \omega, j)$ is the sum of the changes in relative price of equipment, common for all workers (the "shift"), interacted with the sum of the initial shares of workers allocated to that equipment in all occupations (the "share").

The moment condition which identifies my parameter η is thus that the error term $\ln \tilde{Q}(\lambda, \lambda_1, j, j_1, t)^\eta$ is uncorrelated with the instrument $\tilde{\mathcal{X}}_\eta(\lambda, \lambda_1, j, j_1, t)$:

$$\mathbb{E}\left[\Big(\ln \tilde{l}(\lambda, \lambda_1, j, j_1, t) - \eta \ln \tilde{w}(\lambda, \lambda_1, j, j_1, t)\Big) \times \ln \tilde{\mathcal{X}}_\eta(\lambda, \lambda_1, j, j_1)\right] = 0 \tag{2.33}$$

The relevance of the instrument is highlighted by the model as wages depend on the change in equipment prices (equation (2.23)):

$\frac{\hat{w}(\lambda,j)}{\hat{w}(\lambda,j_1)} = \frac{\hat{T}(\lambda,j)}{\hat{T}(\lambda,j_1)} \left(\frac{\hat{s}(\lambda,j)}{\hat{s}(\lambda,j_1)}\right)^{\frac{1}{\theta}}$ and:

$$\hat{s}(\lambda, j) = \sum_{\kappa,\omega} \frac{\hat{q}(\omega)^\theta}{\hat{q}(\omega_1)^\theta} \frac{\hat{q}(\kappa)^\theta}{\hat{q}(\kappa_1)^\theta} \pi_0(\lambda, \kappa, \omega, j)$$

where the exposure of the initial period is taken instead of time 0 to minimize issue of serial correlation with the error term in equation (2.31).

The exclusion restriction holds provided that the instrument $\tilde{\mathcal{X}}_\eta(\lambda, \lambda_1, j, j_1)$ is uncorrelated with the error term $\tilde{Q}(\lambda, \lambda_1, j, j_1, t)$. That means that any change in the local quality of life which is related to the change in the relative price of computerized equipment must be symmetric across skill groups λ.[15] Moreover, the initial shares $\pi_{1979}(\lambda, \kappa, \omega, j)$ should not be correlated with subsequent change in quality of life by skill groups and settlement types.[16] Last, the change in the relative price of equipment $\frac{\hat{q}(\kappa,t)^\theta}{\hat{q}(\kappa_1,t)^\theta}$ must be unrelated to my local group variables.

2.5.4 Estimating the elasticity of substitution between occupations and the dispersion of idiosyncratic preferences for occupation and equipment

I estimate the elasticity of substitution between occupations and the dispersion of idiosyncratic preferences for occupation and equipment jointly following Burstein et al. (2019).

[15] To illustrate, take a smartphone and the experience of a site it gives by providing readily available information on restaurants. It can be argued that the larger the settlement, the more benefit of such a device.

[16] If I exclude the first two waves (1979, 1986) from my analysis and keep the shares from 1979, problems of serial correlations might be considerably reduced. Indeed, the reunification of West and Ost Germany was likely not be anticipated in 1979 and led to considerable geographical disruption, such as the displacement of the capital city from Bonn to Berlin.

Rearranging equation (2.23) and taking log gives:

$$\ln \hat{w}(\lambda, j) = \underbrace{\ln \hat{q}(\omega_1)\hat{q}(\kappa_1)}_{\text{time effect}} + \frac{1}{\theta} \ln \hat{s}(\lambda, j) + \underbrace{\ln \hat{T}(\lambda, j)}_{\text{unobserved}} \tag{2.34}$$

where the wage $\hat{w}(\lambda, j)$ is directly retrieved from the data and the weighted sum of relative equipment and occupation prices $\hat{s}(\lambda, j)$ is retrieved from the data as explained in Section 2.5.3. The time effect $\ln \hat{q}(\omega_1)\hat{q}(\kappa_1)$ drops when controlling for time dummies. The city-skill group specific productivity term $\hat{T}(\lambda, j)$ is unobserved at the time of the estimation as recovering it requires a value for θ.

Likewise, rearranging equation (2.26) and taking log gives:

$$\ln \hat{\zeta}(\omega) = \underbrace{\ln \hat{E} + (1-\alpha)(1-\rho)\hat{q}(\omega_1)}_{\text{time effect}} + \frac{(1-\alpha)(1-\rho)}{\theta} \ln \frac{\hat{q}(\omega)^\theta}{\hat{q}(\omega_1)^\theta} + \underbrace{\ln \hat{a}(\omega)}_{\text{unobserved}} \tag{2.35}$$

where the labor income in occupation ω $\hat{\zeta}(\omega)$ is directly observable in the data and the change in relative occupation price $\frac{\hat{q}(\omega)^\theta}{\hat{q}(\omega_1)^\theta}$ is constructed as explained in Section 2.5.2. The time effect $\ln \hat{E} + (1-\alpha)(1-\rho)\hat{q}(\omega_1)$ drops when controlling for time dummies. The city-skill group specific productivity term $\hat{a}(\omega)$ is unobserved at the time of the estimation as recovering it requires a value for ρ.

Estimating ω and ρ jointly with equations (2.34) and (2.35) with non-linear least squared will lead to biased estimate as the unobservables are related to the regressors as is the case in Section 2.5.3. Indeed, the relative price of occupation in $\hat{s}(\lambda, j)$ is endogenous to the city-skill specific productivity $\hat{T}(\lambda, j)$ as can be seen through equation (2.26). Moreover, the latter equation also makes clear that the occupation specific productivity change $\hat{a}(\omega)$ is related to the relative price of occupation $\frac{\hat{q}(\omega)^\theta}{\hat{q}(\omega_1)^\theta}$. As above, change in relative price of computerized equipment gives a source of variation for the change in equipment and occupation productivity $\hat{s}(\lambda, j)$ as well as for the change in occupation price $\frac{\hat{q}(\omega)^\theta}{\hat{q}(\omega_1)^\theta}$.

I thus use the 2 following instrument for my regressors in equations (2.34) and (2.35) respectively:

$$\mathcal{X}_\theta(\lambda, j, t) = \sum_\kappa \frac{\hat{q}(\kappa, t)^\theta}{\hat{q}(\kappa_1, t)^\theta} \sum_\omega \pi_{1979}(\lambda, \kappa, \omega, j) = \mathcal{X}_\eta(\lambda, j, t) \tag{2.36}$$

$$\mathcal{X}_\rho(\omega, t) = \sum_\kappa \frac{\hat{q}(\kappa, t)^\theta}{\hat{q}(\kappa_1, t)^\theta} \sum_{\lambda, j} \frac{\hat{L}(\lambda)\hat{\xi}(\lambda, j)\pi_{1979}(\lambda, \kappa, \omega, j)}{\sum_{\lambda', j', \kappa'} \hat{L}(\lambda')\hat{\xi}(\lambda', j')\pi_{1979}(\lambda', \kappa', \omega', j')} \tag{2.37}$$

Equations (2.23) and (2.26) from the model show the relevance of both instruments. The exclusion restrictions hold provided that the change in the relative price of computerized equipment and initial factor allocation are unrelated to the city specific skill group productivity and the occupation specific productivity change.

The two moment conditions which identify ω and ρ jointly are thus:

$$\mathbb{E}\left[\left(\ln \hat{w}(\lambda, j) - \ln \hat{q}(\omega_1)\hat{q}(\kappa_1) - \frac{1}{\theta}\ln \hat{s}(\lambda, j)\right) \times \ln \mathcal{X}_{\theta}(\lambda, t)\right] = 0 \tag{2.38}$$

and

$$\mathbb{E}\left[\left(\ln \hat{\zeta}(\omega) - \ln \hat{E} - (1-\alpha)(1-\rho)\hat{q}(\omega_1) - \tfrac{(1-\alpha)(1-\rho)}{\theta}\ln \tfrac{\hat{q}(\omega)^{\theta}}{\hat{q}(\omega_1)^{\theta}}\right) \times \ln \mathcal{X}_{\rho}(\omega, t)\right] = 0 \tag{2.39}$$

In order to wrap up how the model can be operationalized, Appendix B.9 summarizes all the steps I need to perform in order to have my two counterfactuals of interest.

2.6 Discussing the potential bias of the counterfactuals

Some simplifying hypothesis in the model are empirically problematic. Nevertheless, neglected mechanisms affecting either different settlement types or skill group symmetrically will not affect my counterfactuals as I compare sites and skill groups (see equation (2.19)).

The first hypothesis is that the complementarity between computerized equipment and settlement type is taken as given. The model in Chapter 1 showed that population does affect the complementarity between settlement types and computer capital through agglomeration economies, heterogeneously across skill groups. This effect will be captured in my residual $\frac{\hat{T}(\lambda,j)}{\hat{T}(\lambda',j)}$ in equation (2.23). This effect of computerization will thus be kept constant as I perform my counterfactual, whereas it should not. If large cities loose population relative to the rest of the country following recent technological change which is consistent with Ioannides et al. (2008) my counterfactual will overestimate the effect of computerization on skill sorting. Indeed, the decrease in city size would feed back in a lower complementarity between urban centers and computer capital. If large cities gain population instead the impact of computerization on skill sorting is underestimated.

The second problematic hypothesis is the Cobb-Douglas structure of preferences implying that the share of income allocated to housing is constant across skill groups. Nevertheless, this tends to underestimate the effect of computerization on skill sorting. Indeed, absent computerization the lower burden of housing costs on low-skilled in large cities is underestimated, so that their counter-factual proportion in large cities is underestimated.

2.7 Conclusion

Quantifying the impact of computerization on the choice of high-skilled workers to live and work in large cities matters to assess the pressure of computerization on rents in large cities. This paper has provided a tool for this purpose with a model linking the change in the price of computerized equipment relative to other equipment with the change in the skill composition of cities of different sizes. It also allows to see to what extend labor mobility smooth differentials in purchasing power across cities following computerization. Future work could integrate housing policy into the picture by simulating optimal increase in housing supply depending on policy goals. The model could also be use to forecast future pressure on housing market linked to a further decline in the relative price of computerized equipment and taking into account various predictions for changes in skill supply as well as shifts in the occupation structure.

Conclusion

To conclude, Chapter 1 documents a robust positive correlation between the amount of computer capital allocated to a worker and the population of the agglomeration she lives and works in, conditional on observable characteristics. I find that a German worker with mean characteristics has a probability of 0.24 instead of 0.20 to have an advanced computer use at work in large urban centers relative to the rest of the country. As an advanced computer use such as programming requires more computer power compared to basic uses such as word processing, it entails more computer capital. Likewise, a British worker with mean characteristics has a probability of 0.19 instead of 0.16 to have an advanced computer use at work in large urban centers relative to the rest of the country. Occupations not intensive in computer capital are also concerned by the phenomenon. In this case, the variation in computer capital across city size is mainly driven by a larger probability to work with a computer than not. Regarding occupations intensive in computer capital, the variation is mainly driven by a larger probability to have an advanced computer use as almost every worker uses a computer.

The mechanism explaining the city size - computer capital complementarity highlighted in the model in Chapter 1 relies on complex tasks. As complex tasks are complementary with both computer capital and city size through knowledge spillovers, the incentive for firms to allocate computer capital to a given worker increases with city size. This mechanism is supported by some empirical evidence as the computer capital - city size complementarity is driven to zero as I control for direct measures of tasks' complexity at work. High-skilled workers though are more productive at complex tasks and thus spend more time on them relative to simple tasks. As a consequence, they benefit more than workers with a comparative advantage in simple tasks from the additional productivity in large cities. This is due to the substantial investment in computerized equipment there. They thus disproportionately sort in large cities.

Finally, Chapter 2 offers a framework to quantify the impact of computerization on the increasing sorting of high-skilled workers in large cities. In this exercise, I do not dig into the source of the complementarity between city size and computer capital. It also allows to quantify the response of labor mobility across labor market on wage inequalities as a function of city size.

Appendices

Appendix A

Chapter 1

A.1 Theoretical appendix

A.1.1 Agents optimization problem given the choice of the city

Firm's optimization problem

The representative firm minimizes her costs for a given level of local output. The Lagrangien looks as follow, where the Lagrange multiplier W which is the marginal cost of the firm is equal to the price of the fully tradable final good P due to perfect competition and constant returns to scale in K_Z, K_X, X, Z of the production function. This price is chosen as the numéraire and set to 1:

$$\begin{aligned}\mathcal{L}_A = & \; p_Z \int_0^1 K_{Z,A}(s)ds + p_X \int_0^1 K_{X,A}(s)ds + \int_0^1 w_{z,A}(s) Z_A(s)ds + \int_0^1 w_{x,A}(s) X_A(s)ds \\ & - W \left[Y_A - A \left\{ \int_0^1 \phi_Z(s, L_A) \left[K_{Z,A}(s)^\alpha Z_A(s)^{1-\alpha} \right]^{\frac{\beta-1}{\beta}} + \left[K_{X,A}(s)^\alpha X_A(s)^{1-\alpha} \right]^{\frac{\beta-1}{\beta}} ds \right\}^{\beta/(\beta-1)} \right]\end{aligned}$$

where $\phi_Z(s, L_A)$ satisfies assumptions 1.

Worker's optimization problem

$$\max_{l_{z,A}(s)} w_A(s) = l_{z,A}(s)^\delta w_{z,A}(s) + \left[1 - l_{z,A}(s)\right]^\delta w_{x,A}(s) \tag{A.1}$$

where $z_A(s) = l_{z,A}(s)^\delta$ and $x_A(s) = \left[1 - l_{z,A}(s)\right]^\delta$

Restriction on parameters

$\beta > 1$, the elasticity of substitution between inputs' types and $\delta < 1$ in order to have decreasing returns to tasks.

A.1.2 First order conditions and equilibrium

First order conditions of the firm

Note that variables at the level of the skill group (upper case variables) are substituted for by per worker variables (lower case variables) and the population of the skill type in the city $L_A(s)$ as every worker in the city works for the representative firm: $Z_A(s) \equiv z_A(s)L_A(s)$ and $X_A(s) \equiv x_A(s)L_A(s)$:

FOC computer capital $K_{Z,A}(s)$

$$\frac{\partial \mathcal{L}_A}{\partial K_{Z,A}(s)} = 0$$

$$\Leftrightarrow A^{\frac{\beta-1}{\beta}} Y_A^{1/\beta} \phi_Z(s, L_A) Z_A(s)^{(1-\alpha)\frac{\beta-1}{\beta}} \alpha K_{Z,A}(s)^{\frac{\beta(\alpha-1)-\alpha}{\beta}} = p_Z$$

$$\Leftrightarrow K_{Z,A}(s) = A^{\frac{\beta-1}{\beta(1-\alpha)+\alpha}} Y_A^{\frac{1}{\beta(1-\alpha)+\alpha}} \left[\frac{\alpha \phi_Z(s, L_A)}{p_Z}\right]^{\frac{\beta}{\beta(1-\alpha)+\alpha}} Z_A(s)^{\frac{(1-\alpha)(\beta-1)}{\beta(1-\alpha)+\alpha}} \tag{A.2}$$

The second order condition is satisfied as $\frac{\beta(\alpha-1)-\alpha}{\beta} < 0$.

Expression equation (A.2) per worker gives:

$$k_{Z,A}(s) = A^{\frac{\beta-1}{\beta(1-\alpha)+\alpha}} \left[\frac{Y_A}{L_A(s)}\right]^{\frac{1}{\beta(1-\alpha)+\alpha}} \left[\frac{\alpha \phi_Z(s, L_A)}{p_Z}\right]^{\frac{\beta}{\beta(1-\alpha)+\alpha}} z_A(s)^{\frac{(1-\alpha)(\beta-1)}{\beta(1-\alpha)+\alpha}} \tag{A.3}$$

FOC complex tasks $Z_A(s)$:

$$\frac{\partial \mathcal{L}_A}{\partial Z_A(s)} = 0$$

$$\Leftrightarrow A^{\frac{\beta-1}{\beta}} Y_A^{1/\beta} \phi_Z(s, L_A) K_{Z,A}(s)^{\alpha\frac{\beta-1}{\beta}} (1-\alpha) Z_A(s)^{(1-\alpha)\frac{(\beta-1)-\beta}{\beta}} = w_{Z,A}(s) \tag{A.4}$$

The second order condition is satisfied as $\alpha\frac{\beta-1}{\beta} < 0$

Plugging equation (A.2) in equation (A.4) gives:

$$A^{\frac{\beta-1}{\beta(1-\alpha)+\alpha}} \left[\frac{Y_A}{Z_A(s)}\right]^{\frac{1}{\beta(1-\alpha)+\alpha}} \phi_Z(s, L_A)^{\frac{\beta}{\beta(1-\alpha)+\alpha}} (1-\alpha) \left[\frac{\alpha}{p_Z}\right]^{\frac{\alpha(\beta-1)}{\beta(1-\alpha)+\alpha}} = w_{Z,A}(s) \tag{A.5}$$

By symmetry, the optimal level of simple capital per worker and wage rate for simple tasks are given by, respectively:

$$k_{X,A}(s) = A^{\frac{\beta-1}{\beta(1-\alpha)+\alpha}} \left[\frac{Y_A}{L_A(s)}\right]^{\frac{1}{\beta(1-\alpha)+\alpha}} \left[\frac{\alpha}{p_X}\right]^{\frac{\beta}{\beta(1-\alpha)+\alpha}} x_A(s)^{\frac{(1-\alpha)(\beta-1)}{\beta(1-\alpha)+\alpha}} \tag{A.6}$$

$$A^{\frac{\beta-1}{\beta(1-\alpha)+\alpha}} \left[\frac{Y_A}{X_A(s)}\right]^{\frac{1}{\beta(1-\alpha)+\alpha}} (1-\alpha) \left[\frac{\alpha}{p_X}\right]^{\frac{\alpha(\beta-1)}{\beta(1-\alpha)+\alpha}} = w_{X,A}(s) \tag{A.7}$$

Relative tasks demand

Combining equation (A.5) and equation (A.7) gives the relative tasks demand from the firm:

$$\frac{w_{z,A}(s)}{w_{x,A}(s)} = \phi_Z(s, L_A)^{\frac{\beta}{\beta(1-\alpha)+\alpha}} \left[\frac{p_X}{p_Z}\right]^{\frac{\alpha(\beta-1)}{\beta(1-\alpha)+\alpha}} \left[\frac{x_A(s)}{z_A(s)}\right]^{\frac{1}{\beta(1-\alpha)+\alpha}} \tag{A.8}$$

Relative tasks supply

Then the optimality condition of workers (maximizing equation (1.4)) gives the relative supply for tasks as a function of the wage rates for complex $w_z(s, A)$ and simple tasks $w_x(s, A)$ (the second order conditions are satisfied given that $\delta < 1$) :

$$\frac{w_{z,A}(s)}{w_{x,A}(s)} = \left[\frac{z_A(s)}{x_A(s)}\right]^{\frac{1-\delta}{\delta}} \tag{A.9}$$

Equating supply (equation (A.9)) and demand (equation (A.8)) implicitly defines the optimal share of time allocated to complex tasks as function of exogenous variables given city population:

Equating demand and supply gives

$$\phi_Z(s, L_A)^{\frac{\beta}{\beta(1-\alpha)+\alpha}} \left[\frac{p_X}{p_Z}\right]^{\frac{\alpha(\beta-1)}{\beta(1-\alpha)+\alpha}} = \left[\frac{l_{z,A}(s)}{1-l_{z,A}(s)}\right]^{\frac{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}{\beta(1-\alpha)+\alpha}}$$

where $z_A(s) = l_{z,A}(s)^{\delta}$ and $x_A(s) = (1 - l_{z,A}(s))^{\delta}$

$$\Leftrightarrow B_{Z,A}(s) = \frac{l_{z,A}(s)}{1 - l_{z,A}(s)}$$

where $B_{Z,A}(s) = \phi_Z(s, L_A)^{\frac{\beta}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}} \left[\frac{p_X}{p_Z}\right]^{\frac{\alpha(\beta-1)}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}}$

$$\Leftrightarrow B_{Z,A}(s) = l_{z,A}(s)\left(1 + B_{Z,A}(s)\right)$$

$$\Leftrightarrow l_{z,A}(s) = \frac{B_{Z,A}(s)}{1 + B_{Z,A}(s)} \tag{A.10}$$

Because of decreasing returns to tasks ($\delta < 1$), every worker spends a positive amount of time on each type of tasks. I impose $\frac{\partial^2 \phi_Z(s,L_A)}{\partial^2 L_A} < 0$ and $\frac{\partial^2 \phi_Z(s,L_A)}{\partial^2 s} < 0$ so that $B_{Z,A}(s)$ remains bounded for all possible s and L_A. Note that the optimal amount of time allocated to complex tasks is independent from the local skill labor supply $L_A(s)$, and that $a \equiv \frac{\beta}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta} > 1$.

Proof of Lemma 1 part (i):

The behavior of the share of working time allocated to complex tasks (equation (A.10)) with s and L_A in percent depends on assumption 1 part (iv): $\frac{\partial^2 \ln \phi_{z,A}(s)}{\partial L_A \partial s}$ is either equal to zero or increasing.

First of all, the log of the optimal amount of working time to complex tasks is equal to:

$$\ln l_{z,A}(s) = a \ln \phi_Z(s, L_A) + \ln \left[\frac{p_X}{p_Z}\right]^{\frac{\alpha(\beta-1)}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}} - \ln\left(1 + \phi_Z(s, L_A)^a \left[\frac{p_X}{p_Z}\right]^{\frac{\alpha(\beta-1)}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}}\right) \tag{A.11}$$

The derivative of this expression with respect to L_A is equal to:

$$\frac{\partial \ln l_{z,A}(s)}{\partial L_A} = \frac{a}{\phi_Z(s, L_A)} \frac{\partial \phi_Z(s, L_A)}{\partial L_A} - \frac{B_{Z,A}(s)}{(1 + B_{Z,A}(s))} \frac{a}{\phi_Z(s, L_A)} \frac{\partial \phi_Z(s, L_A)}{\partial L_A}$$

$$= a \frac{\partial \ln \phi_Z(s, L_A)}{\partial L_A} \left[\frac{1}{(1 + B_{Z,A}(s))}\right] > 0 \tag{A.12}$$

where $a = \frac{\beta}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}$

Next, this derivative w.r.t. s gives:

$$\frac{\partial^2 \ln l_{z,A}(s)}{\partial L_A \partial s} = \frac{a}{(1+B_{Z,A}(s))}\frac{\partial^2 \ln \phi_Z(s,L_A)}{\partial L_A \partial s} - a\frac{\partial \ln \phi_Z(s,L_A)}{\partial L_A}\frac{\partial \ln \phi_Z(s,L_A)}{\partial s}\frac{B_{Z,A}(s)}{(1+B_{Z,A}(s))^2} \tag{A.13}$$

This expression is negative if $\frac{\partial^2 \log \phi_Z(s,L_A)}{\partial L \partial s} = 0$ and theoretically ambiguous if $\frac{\partial^2 \log \phi_Z(s,L_A)}{\partial L \partial s} > 0$. This latter result relates to two opposite effects. The log supermodularity of $\phi_Z(s,L_A)$ in s and L_A implies that an increase in city size will disproportionately impact the allocation of time to complex tasks for the high-skilled population. However, this population spends a larger share of time on complex tasks regardless of location. The effect of decreasing returns to tasks ($\delta < 1$) thus becomes more salient for them. In the first case where $\frac{\partial^2 \log \phi_Z(s,L_A)}{\partial L \partial s} = 0$, only the effect related to the decreasing returns to tasks is present and the percentage increase in time allocated to complex tasks with city size is smaller for high-skilled workers relative to low-skilled ones.

Proof of Lemma 1 part (ii):

The log of simple tasks is $-\ln(1+B_{Z,A}(s))$, its derivative w.r.t. L_A is equal to:

$$\frac{\partial \ln(1-l_{z,A}(s))}{\partial L_A} = -\frac{aB_{Z,A}(s)}{(1+B_{Z,A}(s))}\frac{\partial \ln \phi_Z(s,L_A)}{\partial L_A}$$

where $\frac{B_{Z,A}(s)}{\left(1+B_{Z,A}(s)\right)} = l_{z,A}(s)$. Because $l_{z,A}(s)$ is increasing in s and $\frac{\partial^2 \log \phi_Z(s,L_A)}{\partial L \partial s} \geq 0$, $\frac{\partial^2 \ln(1-l_{z,A}(s))}{\partial L_A \partial s} < 0 \ \forall \ s$. That is, the effect becomes even more negative for high-skilled workers. The result is not symmetric to Lemma 1 part (i) as it focuses on percentage increase of share of time allocated to simple tasks (high-skilled workers have overall a lower share of time allocated to simple tasks). Focusing on the share of time allocated to complex and simple tasks instead of their log shows perfect symmetry between the two: the change in time allocated to complex tasks goes to simple ones.

Rewriting wage and capital given city population

The wage of the worker $w_A(s)$ (equation (A.1)) can then be rewritten as:

$$w_A(s) = A^{\frac{\beta-1}{\beta(1-\alpha)+\alpha}} \left[\frac{Y_A}{L_A(s)}\right]^{\frac{1}{\beta(1-\alpha)+\alpha}} (1-\alpha) \left[\frac{\alpha}{p_X}\right]^{\frac{\alpha(\beta-1)}{\beta(1-\alpha)+\alpha}}$$
$$\left[l_{z,A}(s)^{\frac{\delta(1-\alpha)(\beta-1)}{\beta(1-\alpha)+\alpha}} \phi_Z(s, L_A)^{\frac{\beta}{\beta(1-\alpha)+\alpha}} \left[\frac{p_X}{p_Z}\right]^{\frac{\alpha(\beta-1)}{\beta(1-\alpha)+\alpha}} + (1-l_{z,A}(s))^{\frac{\delta(1-\alpha)(\beta-1)}{\beta(1-\alpha)+\alpha}} \right] \tag{A.14}$$

Plugging the optimal amount of time allocated to complex tasks (equation (A.10)) back in the wage of the workers gives:

$$w_A(s) = A^{\frac{\beta-1}{\beta(1-\alpha)+\alpha}} \left[\frac{Y_A}{L_A(s)}\right]^{\frac{1}{\beta(1-\alpha)+\alpha}} (1-\alpha) \left[\frac{\alpha}{p_X}\right]^{\frac{\alpha(\beta-1)}{\beta(1-\alpha)+\alpha}}$$
$$\left[\phi_Z(s, L_A)^{\frac{\beta}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}} \left[\frac{p_X}{p_Z}\right]^{\frac{\alpha(\beta-1)}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}} + 1 \right]^{1-\frac{\delta(1-\alpha)(\beta-1)}{\beta(1-\alpha)+\alpha}} \tag{A.15}$$

$$= A^{\frac{\beta-1}{\beta(1-\alpha)+\alpha}} \left[\frac{Y_A}{L_A(s)}\right]^{\frac{1}{\beta(1-\alpha)+\alpha}} (1-\alpha) \left[\frac{\alpha}{p_X}\right]^{\frac{\alpha(\beta-1)}{\beta(1-\alpha)+\alpha}} [1 + B_{Z,A}(s)]^{\frac{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}{\beta(1-\alpha)+\alpha}} \tag{A.16}$$

The optimal amount of computer capital then becomes (per worker) (plugging equation (A.10) in equation (A.3), with $z_A(s) = l_{z,A}(s)^{\delta}$):

$$k_{Z,A}(s) = A^{\frac{\beta-1}{\beta(1-\alpha)+\alpha}} \left[\frac{Y_A \alpha^{\beta}}{L_A(s)}\right]^{\frac{1}{\beta(1-\alpha)+\alpha}} \phi_Z(s, L_A)^{\frac{\beta}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}} p_Z^{\frac{\delta(1-\alpha)(\beta-1)-\beta}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}}$$
$$p_X^{\frac{\delta(1-\alpha)(\beta-1)}{\beta(1-\alpha)+\alpha} \frac{\alpha(\beta-1)}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}} [1 + B_{Z,A}(s)]^{-\frac{\delta(1-\alpha)(\beta-1)}{\beta(1-\alpha)+\alpha}} \tag{A.17}$$

Similarly, the optimal amount of simple capital becomes:

$$k_{X,A}(s) = A^{\frac{\beta-1}{\beta(1-\alpha)+\alpha}} \left[\frac{Y_A \alpha^{\beta}}{L_A(s)}\right]^{\frac{1}{\beta(1-\alpha)+\alpha}} p_X^{\frac{-\beta}{\beta(1-\alpha)+\alpha}} [1 + B_{Z,A}(s)]^{-\frac{\delta(1-\alpha)(\beta-1)}{\beta(1-\alpha)+\alpha}} \tag{A.18}$$

The ratio between simple and complex capital per worker thus is equal to:

$$\frac{k_{Z,A}(s)}{k_{X,A}(s)} = \phi_Z(s, L_A)^{\frac{\beta}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}} \left[\frac{p_X}{p_Z}\right]^{\frac{-\delta(1-\alpha)(\beta-1)+\beta}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}} \tag{A.19}$$

Recovering Y_A

Substituting all endogenous variables derived above in the equation for local output Y_A and taking in front of the integral everything common in simple and complex capital: $A^{\alpha \frac{\beta-1}{\beta(1-\alpha)+\alpha}} \left(Y_A \alpha^{\beta}\right)^{\frac{\alpha}{\beta(1-\alpha)+\alpha}}$ simplifies to:

$$Y_A = A \left\{ \int_0^1 \phi_{Z,A}(s) \left[K_{Z,A}(s)^{\alpha} Z_A(s)^{1-\alpha}\right]^{\frac{\beta-1}{\beta}} + \left[K_{X,A}(s)^{\alpha} X_A(s)^{1-\alpha}\right]^{\frac{\beta-1}{\beta}} ds \right\}^{\beta/(\beta-1)}$$

$$= A^{\frac{\beta}{\beta(1-\alpha)+\alpha}} \left(Y_A \alpha^{\beta}\right)^{\frac{\alpha}{\beta(1-\alpha)+\alpha}} p_X^{\frac{-\beta\alpha}{\beta(1-\alpha)+\alpha}} \left\{ \int_0^1 L_A(s)^{\frac{(\beta-1)(1-\alpha)}{\beta(1-\alpha)+\alpha}} \left[1 + B_{Z,A}(s)\right]^{\frac{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}{\beta(1-\alpha)+\alpha}} ds \right\}^{\beta/(\beta-1)}$$

$$Y_A = Y_A^{\frac{\alpha}{\beta(1-\alpha)+\alpha}} C$$

$$Y_A = C^{\frac{\beta(1-\alpha)+\alpha}{\beta(1-\alpha)}}$$

$$Y_A = A^{\frac{1}{1-\alpha}} \left[\frac{\alpha}{p_X}\right]^{\frac{\alpha}{1-\alpha}} \left\{ \int_0^1 L_A(s)^{\frac{(\beta-1)(1-\alpha)}{\beta(1-\alpha)+\alpha}} \left[1 + B_{Z,A}(s)\right]^{\frac{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}{\beta(1-\alpha)+\alpha}} ds \right\}^{\frac{\beta(1-\alpha)+\alpha}{(1-\alpha)(\beta-1)}} \tag{A.20}$$

A.1.3 Properties of the wage of workers given city population

In order to prove Lemma 2 part (i), (ii) and (iii), let us rewrite equation (A.14) substituting for the expression for total city output Y_A (equation (A.20)):

$$\begin{aligned} w_A(s) = & A^{\frac{1}{1-\alpha} + \frac{\beta-1}{\beta(1-\alpha)+\alpha}} \left[\frac{\alpha}{p_X}\right]^{\frac{\alpha}{1-\alpha} + \frac{\alpha(\beta-1)}{\beta(1-\alpha)+\alpha}} L_A(s)^{\frac{-1}{\beta(1-\alpha)+\alpha}} (1-\alpha) \left[1 + B_{Z,A}(s)\right]^{\frac{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}{\beta(1-\alpha)+\alpha}} \\ & \left\{ \int_0^1 L_A(s)^{\frac{(\beta-1)(1-\alpha)}{\beta(1-\alpha)+\alpha}} \left[1 + B_{Z,A}(s)\right]^{\frac{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}{\beta(1-\alpha)+\alpha}} ds \right\}^{\frac{\beta(1-\alpha)+\alpha}{(1-\alpha)(\beta-1)}} \end{aligned} \tag{A.21}$$

where $B_{Z,A}(s) = \phi_Z(s, L_A)^{\frac{\beta}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}} \left[\frac{p_X}{p_Z}\right]^{\frac{\alpha(\beta-1)}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}}$.

Proofs:

Lemma 2 part (i)

The wage of a worker with skill s - all else equal, is log-supermodular in s and L given the average skill distribution of the city for all parameter values: $\frac{\partial^2 \log w(s,A)}{\partial s \partial L} > 0$:

Given the skill distribution, the only term depending on s is $[1+B_{Z,A}(s)]^{\frac{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}{\beta(1-\alpha)+\alpha}} \equiv n_A(s)$, where $B_{Z,A}(s) = \phi_Z(s,L_A)^{\frac{\beta}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}} \left[\frac{p_X}{p_Z}\right]^{\frac{\alpha(\beta-1)}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}}$. Thus, if this term is log-supermodular in s and L, so will be the wage:

Taking the derivative of $\ln n_A(s) = \frac{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}{\beta(1-\alpha)+\alpha} \ln\left[1+B_{Z,A}(s)\right]$ w.r.t L_A:

$$\frac{\partial \ln n_A(s)}{\partial L_A} = \left[1+B_{Z,A}(s)\right]^{-1} B_{Z,A}(s) \frac{\beta}{\beta(1-\alpha)+\alpha} \frac{\partial \log \phi_Z(s,L_A)}{\partial L_A} > 0 \tag{A.22}$$

The derivative of the derivative w.r.t. s:

$$\begin{aligned}\frac{\partial^2 \ln n_A(s)}{\partial L_A \partial s} &= \left[1+B_{Z,A}(s)\right]^{-1} B_{Z,A}(s) \frac{\beta}{\beta(1-\alpha)+\alpha} \frac{\partial^2 \log \phi_Z(s,L_A)}{\partial L_A \partial s} \\ &+ \left[1+B_{Z,A}(s)\right]^{-1} B_{Z,A}(s) \frac{a\beta}{\beta(1-\alpha)+\alpha} \frac{\partial \log \phi_Z(s,L_A)}{\partial L_A} \frac{\partial \log \phi_Z(s,L_A)}{\partial s} \\ &- \left[1+B_{Z,A}(s)\right]^{-2} B_{Z,A}(s) \frac{a\beta}{\beta(1-\alpha)+\alpha} \frac{\partial \log \phi_Z(s,L_A)}{\partial L_A} \frac{\partial \log \phi_Z(s,L_A)}{\partial s}\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 \ln n_A(s)}{\partial L_A \partial s} &= \left[1+B_{Z,A}(s)\right]^{-1} B_{Z,A}(s) \frac{\beta}{\beta(1-\alpha)+\alpha} \frac{\partial^2 \log \phi_Z(s,L_A)}{\partial L_A \partial s} \\ &+ \frac{B_{Z,A}(s)^2}{\left[1+B_{Z,A}(s)\right]^2} \frac{a\beta}{\beta(1-\alpha)+\alpha} \frac{\partial \log \phi_Z(s,L_A)}{\partial L_A} \frac{\partial \log \phi_Z(s,L_A)}{\partial s} > 0\end{aligned} \tag{A.23}$$

where $a = \frac{\beta}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}$ and $B_{Z,A}(s) = \phi_Z(s,L_A)^{\frac{\beta}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}} \left[\frac{p_X}{p_Z}\right]^{\frac{\alpha(\beta-1)}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}}$.

Lemma 2 part (ii)

The wage of a worker with skill s - all else equal, the log supermodularity of the wage of the worker in s and L increases as the relative price of computerized equipment p_Z relative to simple equipment p_X decreases, for all parameter values. Indeed, as $l_{z,A}(s) = \dfrac{B_{Z,A}(s)}{1+B_{Z,A}(s)}$ is increasing in p_X/p_Z, equation (A.23) is increasing in it.

A.1.4 The rent given population

Equation (1.2) $r_A = (1-\mu)\frac{L_A}{H}\bar{w}_A$ can be rewritten using equation (A.21) as follows:

$$r_A = \frac{(1-\mu)}{\overline{H}} A^{\frac{1}{1-\alpha}+\frac{\beta-1}{\beta(1-\alpha)+\alpha}} \left[\frac{\alpha}{p_X}\right]^{\frac{\alpha}{1-\alpha}+\frac{\alpha(\beta-1)}{\beta(1-\alpha)+\alpha}} (1-\alpha)$$
$$\left\{\int_0^1 L_A(s)^{\frac{(\beta-1)(1-\alpha)}{\beta(1-\alpha)+\alpha}} \left[1+B_{Z,A}(s)\right]^{\frac{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}{\beta(1-\alpha)+\alpha}} ds\right\}^{\frac{(1-\alpha)(2\beta-1)+\alpha}{(1-\alpha)(\beta-1)}} \tag{A.24}$$

$$r_A^{\mu-1} = \left(\frac{(1-\mu)}{\overline{H}} A^{\frac{1}{1-\alpha}+\frac{\beta-1}{\beta(1-\alpha)+\alpha}} \left[\frac{\alpha}{p_X}\right]^{\frac{\alpha}{1-\alpha}+\frac{\alpha(\beta-1)}{\beta(1-\alpha)+\alpha}} (1-\alpha)\right)^{\mu-1}$$
$$\left\{\int_0^1 L_A(s)^{\frac{(\beta-1)(1-\alpha)}{\beta(1-\alpha)+\alpha}} \left[1+B_{Z,A}(s)\right]^{\frac{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}{\beta(1-\alpha)+\alpha}} ds\right\}^{\frac{(1-\alpha)(2\beta-1)+\alpha}{(1-\alpha)(\beta-1)}(\mu-1)}$$

A.1.5 Indirect utility, existence and stability of spatial equilibrium

To begin with, let us rewrite the indirect utility in equation (1.1) using the expressions for rent in equation (A.24), and for wage in equation (A.21).

$$V_A(s) = \left(\frac{(1-\mu)}{\overline{H}}\right)^{\mu-1} \left(A^{\frac{1}{1-\alpha}+\frac{\beta-1}{\beta(1-\alpha)+\alpha}} \left[\frac{\alpha}{p_X}\right]^{\frac{\alpha}{1-\alpha}+\frac{\alpha(\beta-1)}{\beta(1-\alpha)+\alpha}} (1-\alpha)\right)^{\mu}$$
$$L_A(s)^{\frac{-1}{\beta(1-\alpha)+\alpha}} \left[1+B_{Z,A}(s)\right]^{\frac{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}{\beta(1-\alpha)+\alpha}}$$
$$\left\{\int_0^1 L_A(s)^{\frac{(\beta-1)(1-\alpha)}{\beta(1-\alpha)+\alpha}} \left[1+B_{Z,A}(s)\right]^{\frac{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}{\beta(1-\alpha)+\alpha}} ds\right\}^{\frac{\mu[(1-\alpha)(2\beta-1)+\alpha]-(1-\alpha)(\beta-1)}{(1-\alpha)(\beta-1)}} \tag{A.25}$$

where $B_{Z,A}(s) = \phi_Z(s, L_A)^{\frac{\beta}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}} \left[\frac{p_X}{p_Z}\right]^{\frac{\alpha(\beta-1)}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}}$.

To arrive to that equation, factor out in front of the integral of the average wage every term which does not depend on s and collect terms. The **no-black hole condition** which is necessary for the only equilibrium not to be one in which every worker agglomerates in a single city is $\frac{\mu[(1-\alpha)(2\beta-1)+\alpha]-(1-\alpha)(\beta-1)}{(1-\alpha)(\beta-1)} < 0 \leftrightarrow \mu < \left[2+\frac{1}{(1-\alpha)(\beta-1)}\right]^{-1} \in \left(0, \frac{1}{2}\right)$ or $1-\mu > \frac{1}{2}$. That is, the preferences of workers for the final good must not be too large.

$$\leftrightarrow \mu < \frac{(1-\alpha)(\beta-1)}{(1-\alpha)(2\beta-1)+\alpha} \tag{A.26}$$

In a next step, in order to have an explicit system of differential equations, I isolate $L_A(s)$ in equation (A.25) and replace the indirect utility $V_A(s)$ by the constant utility across all sites $V_E(s)$ compatible with a spatial equilibrium where all sites are populated:

$$L_A(s) = V_E(s)^{-\beta(1-\alpha)-\alpha} \left[1 + B_{Z,A}(s)\right]^{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta} C_A^{\beta(1-\alpha)+\alpha} \tag{A.27}$$

where $B_{Z,A}(s) = \phi_Z(s, L_A)^{\frac{\beta}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}} \left[\frac{p_X}{p_Z}\right]^{\frac{\alpha(\beta-1)}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}}$
and $C_A = \left(\frac{(1-\mu)}{\overline{H}}\right)^{\mu-1} \left(A^{\frac{1}{1-\alpha}+\frac{\beta-1}{\beta(1-\alpha)+\alpha}} \left[\frac{\alpha}{p_X}\right]^{\frac{\alpha}{1-\alpha}+\frac{\alpha(\beta-1)}{\beta(1-\alpha)+\alpha}} (1-\alpha)\right)^{\mu}$
$\left\{\int_0^1 L_A(s)^{\frac{(\beta-1)(1-\alpha)}{\beta(1-\alpha)+\alpha}} \left[1 + B_{Z,A}(s)\right]^{\frac{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}{\beta(1-\alpha)+\alpha}} ds\right\}^{\frac{\mu[(1-\alpha)(2\beta-1)+\alpha]-(1-\alpha)(\beta-1)}{(1-\alpha)(\beta-1)}}$.

The population clearing condition $L(s) = \sum_{A\in\mathcal{A}} L_A(s)$ allows to recover the value for $V_E(s)$ as a function of city sizes:

$$V_E(s)^{-\beta(1-\alpha)-\alpha} = \frac{L(s)}{\sum_{A\in\mathcal{A}} \left[1 + B_{Z,A}(s)\right]^{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta} C_A^{\beta(1-\alpha)+\alpha}} \tag{A.28}$$

$$V_E(s) = L(s)^{\frac{-1}{\beta(1-\alpha)+\alpha}} \left(\sum_{A\in\mathcal{A}} \left[1 + B_{Z,A}(s)\right]^{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta} C_A^{\beta(1-\alpha)+\alpha}\right)^{\frac{1}{\beta(1-\alpha)+\alpha}} \tag{A.29}$$

Back in equation (A.27):

$$L_A(s) = \frac{L(s) \left[1 + B_{Z,A}(s)\right]^{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta} C_A^{\beta(1-\alpha)+\alpha}}{\sum_{A\in\mathcal{A}} \left[1 + B_{Z,A}(s)\right]^{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta} C_A^{\beta(1-\alpha)+\alpha}}$$

Everything which does not depend on the location A simplifies, where the endogenous function $L_A(s)$ is in bold in the equation:

$$\mathbf{L_A}(\mathbf{s}) = \frac{L(s) \left[1 + B_{Z,A}(s)\right]^{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta} D_A^{\beta(1-\alpha)+\alpha}}{\sum_{A\in\mathcal{A}} \left[1 + B_{Z,A}(s)\right]^{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta} D_A^{\beta(1-\alpha)+\alpha}} \tag{A.30}$$

with
$D_A = A^{\mu\left[\frac{1}{1-\alpha}+\frac{\beta-1}{\beta(1-\alpha)+\alpha}\right]} \left\{\int_0^1 \mathbf{L_A}(\mathbf{s})^{\frac{(\beta-1)(1-\alpha)}{\beta(1-\alpha)+\alpha}} \left[1 + B_{Z,A}(s)\right]^{\frac{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}{\beta(1-\alpha)+\alpha}} ds\right\}^{\frac{\mu[(1-\alpha)(2\beta-1)+\alpha]-(1-\alpha)(\beta-1)}{(1-\alpha)(\beta-1)}}$

and $B_{Z,A}(s) = \phi_Z(s, \int_0^1 \mathbf{L_A}(\mathbf{s}) ds)^{\frac{\beta}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}} \left[\frac{p_X}{p_Z}\right]^{\frac{\alpha(\beta-1)}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}}$.

Let $U \subset \mathbb{R}^n$ be the set of city sizes, where n is the number of discrete cities.

$$U := \left\{ L_A(s) \in [0, L(s)] \text{ continuous on } I = [0,1]\,, \sum_{A \in \mathcal{A}} L_A(s) = L(s) \right\}$$

U is convex and bounded by the total population in the economy $L(s)$, where the bounds pertain to the domain, so also closed.

Equation (A.30) is a system of ordinary differential equations $f : U \times [0,1] \to \mathbb{R}^n$ mapping the population of each city $L_A(s)$ to the skill $s \in [0,1]$.[1] The domain of the RHS of equation (A.30) is also between 0 and $L(s)$, including the bounds. D_A diverges to infinity when $L_A(s) = 0$ due to the no black-hole condition. Nevertheless, the RHS of equation (A.30) is also defined in this case as D_A also appears in the denominator.

I look for a fixed point of the equation $(TL_A(s))$ as defined in (A.30). That is I want to find $L_A(s) \in U$ with $TL_A(s) = L_A(s)$. Note that $T : U \to U$. T is continuous since I assume $L(s)$ to be continuous in s and $\phi_Z(s, L_A)$ is twice differentiable in s and L_A. T has a fixed point as a consequence of the Schauder-Tychonoff Theorem.[2]

I have thus proved that a solution exists, where the solution is for each sites between being totally empty and receiving the whole population to be allocated. I am though not interested in a corner solution where $L_A(s)$ is either 0 or $L(s)$. Equation (A.30) shows that this is not a solution. Indeed, utility would be zero in unpopulated sites and strictly positive elsewhere. To see this, I take the case where population is 0 in one site denoted by 1 and strictly positive elsewhere. Applying the function on the RHS of equation (A.30) gives $L_1(s) = L(s)$ which contradicts $L_1(s) = 0$. The reasoning translates to any number of sites being unpopulated, where applying the RHS of equation (A.30) each time results in strictly positive local population.

Computation: I rewrite the equilibrium population from equation (A.30) for this site as follow:

$$L_1(s) = \frac{L(s)\left[1 + B_{Z,1}(s)\right]^{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}}{\frac{1}{D_1^{\beta(1-\alpha)+\alpha}} \sum_{A \in \mathcal{A}} \left[1 + B_{Z,A}(s)\right]^{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta} D_A^{\beta(1-\alpha)+\alpha}}$$

When the population in the site number 1 tends to zero whereas it remains positive for all other sites, D_1 will tend to infinity because of the no black-hole condition. All terms in the sum in the denominator will thus tend to zero except for site 1:

[1]Equation (A.30) differs from a standard initial value problem $y'(x) = f(x, y(x)), y(x_0) = y_0$ because the integral in the function f does not depend on x: $y'(x) = f(x, \int_0^1 y'(x)dx)$. Also in the integral over $x = [0,1]$, $y'(x)$ appears weighted in D_A

[2]If T: $X \to X$ is continuous and if $A \subset X$ is a convex compact subset of the normed linear space X and $T(A) \subset A$, then T has a fixed point. In my case, the subset is my original set U.

$$L_1(s) = \frac{L(s)\left[1 + B_{Z,1}(s)\right]^{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}}{\left[1 + B_{Z,1}(s)\right]^{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}} = L(s)$$

Intuition:

Consider the convergence of a sequence of functions $L_A(s)^m$ when the initial chosen function is the same population in the n sites:

$$L_A(s)^m = \begin{cases} \frac{L(s)}{n} & m = 0 \\ f_A(L_A(s)^{m-1}) & m > 0 \end{cases}$$

where the function $f_A(.)$ is given by the right hand side of equation (A.30).

Consider the symmetric allocation. The utility of workers will be larger in sites with a larger exogenous productivity A. Increasing population there rebalances utility across sites as $\frac{\partial V_A}{\partial L_A} < 0$ for the majority of the population (see condition in Appendix A.1.6). Because of the Inada condition ($\lim_{L_A \to 0} \frac{\partial \phi}{\partial L_A} \to \infty$, $\lim_{L_A \to L} \frac{\partial \phi}{\partial L_A} \to 0$), a given increase in population leads to a stronger decrease in utility where population is high. Thus, higher population in high-A sites can balance average utility over sites. Now, agglomeration economies are skill-biased. To compensate, the proportion of high-skilled workers is increased with A. This is feasible as $\lim_{g_A(s) \to 0} \frac{\partial V_A(s)}{\partial g_A(s)} \to \infty$, $\lim_{g_A(s) \to 1} \frac{\partial V_A(s)}{\partial g_A(s)} \to 0$ as all skill types are needed to produce locally. The change in population necessary for a given change in utility thus gets smaller as the proportion of workers with the given skill gets smaller. Equation (A.30) does indeed allocate more population in sites where utility is disproportionately high when iterating the sequence. When this is the case, the numerator of the RHS of equation (A.30) is larger than its denominator.

I now discuss the case where utility is increasing with city size for some top skilled worker for low range of city size. For them when starting from the symmetric allocation, decreasing city size in low-A might decrease their utility there and increase their utility in high-A if the symmetric allocation for them is in the range of sizes where $\frac{V_A}{L_A} > 0$, further strengthening the difference in utility across sites. Nevertheless, the change in skill density will be able to rebalance this effect. The more numerous the people for which it is the case, the less be will the difference in city size across sites and the more the difference in the skill distribution across them.

Stability of the solution:

I have defined stability as follow. The utility of any workers with skill s must be decreasing or be unaffected if a new inhabitant comes to the city in equilibrium whatever the skill $\check{s}$ of the new comer:

$$\frac{\partial V_A(s)}{\partial L_{E,A}(\check{s})} \leq 0 \ \forall \check{s}; \forall s.$$

For this condition to be true, I need $\frac{V_A}{L_A} < 0$ for the whole range of the skill distribution. I need additional restriction on the variation of ability to perform complex tasks $\phi_Z(s, L_A)$. If the distribution of ability is too unequal, top performer will benefit from a new comer, as average wage on which housing costs depend will not weight much for them. But this would not lead population to diverge to agglomerating there, because because top skilled workers would come to that city rebalancing there utility there.

A.1.6 Proof of proposition 1

Proposition 1 part (i)

The population increases with the exogenous productivity of the site: $\frac{\partial L_E(A)}{\partial A} > 0$:

The total derivative of the utility of workers across cities (equation (A.25)) must be equalized by the spatial equilibrium condition:

$$dV = \frac{\partial V_A(s)}{\partial A} dA + \frac{\partial V_A(s)}{\partial g_A(s)} dg_A(s) + \frac{\partial V_A(s)}{\partial L_A} dL_A = 0 \tag{A.31}$$

where $g_A(s) = \frac{L_A(s)}{L_A}$ is the proportion of workers with skill s, $L_A = \int_0^1 L_A(s) ds$ and A exogenously differs across sites.

Let us rewrite equation (A.25) as a function of L_A and $g_A(s)$ in order to analyze the behavior of indirect utility separately for these two variable, as well as a function of the average wage in the city (worker with skill $\tilde{s}$):

$$\begin{aligned} V_A(s) = & \left(\frac{(1-\mu)}{\overline{H}}\right)^{\mu-1} \left(A^{\frac{1}{1-\alpha}+\frac{\beta-1}{\beta(1-\alpha)+\alpha}} \left[\frac{\alpha}{p_X}\right]^{\frac{\alpha}{1-\alpha}+\frac{\alpha(\beta-1)}{\beta(1-\alpha)+\alpha}} (1-\alpha) \right)^{\mu} \\ & g_A(s)^{\frac{-1}{\beta(1-\alpha)+\alpha}} \left[1 + B_{Z,A}(s)\right]^{\frac{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}{\beta(1-\alpha)+\alpha}} L_A^{\tilde{\mu}-\frac{1}{\beta(1-\alpha)+\alpha}} \\ & \left\{ g_A(\tilde{s})^{\frac{-1}{\beta(1-\alpha)+\alpha}} \left[1 + B_{Z,A}(\tilde{s})\right]^{\frac{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}{\beta(1-\alpha)+\alpha}} \right\}^{\tilde{\mu}} \end{aligned} \tag{A.32}$$

where $B_{Z,A}(s) = \phi_Z(s, L_A)^{\frac{\beta}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}} \left[\frac{p_X}{p_Z}\right]^{\frac{\alpha(\beta-1)}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}}$, $\tilde{\mu} = \frac{\mu[(1-\alpha)(2\beta-1)+\alpha]-(1-\alpha)(\beta-1)}{(1-\alpha)(\beta-1)} < 0$ and $\int_0^1 g_A(s)^{\frac{-1}{\beta(1-\alpha)+\alpha}} \left[1 + B_{Z,A}(s)\right]^{\frac{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}{\beta(1-\alpha)+\alpha}} ds = g_A(\tilde{s})^{\frac{-1}{\beta(1-\alpha)+\alpha}} \left[1 + B_{Z,A}(\tilde{s})\right]^{\frac{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}{\beta(1-\alpha)+\alpha}}$

What happens when dA is positive, i.e. as the exogenous productivity of the site increases? The wage of workers increases with A for a given population, which translates to the utility: $\frac{\partial V(s,A)}{\partial A} > 0$. The wage and utility of workers a given skill group s when the proportion of that skill group increases : $\frac{\partial V_A(s)}{\partial g_A(s)} < 0$. The change in the proportion of the population of one skill type relative to others cannot have the same sign $\forall s$ by definition, unless it is equal to zero for all workers. Consequently, a positive $dg_A(s)$ cannot balance the positive term $\frac{\partial V_A(s)}{\partial A}dA$ for utility to be equalized **for all skill types**.

Next, I compute $\frac{\partial V_A(s)}{\partial L_A}$:

$$
\begin{aligned}
=&\tilde{C} g_A(s)^{\frac{-1}{\beta(1-\alpha)+\alpha}} \left[1 + B_{Z,A}(s)\right]^{\frac{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}{\beta(1-\alpha)+\alpha}} L_A^{\tilde{\mu}-\frac{1}{\beta(1-\alpha)+\alpha}} \\
&\left\{ g_A(\tilde{s})^{\frac{-1}{\beta(1-\alpha)+\alpha}} \left[1 + B_{Z,A}(\tilde{s})\right]^{\frac{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}{\beta(1-\alpha)+\alpha}} \right\}^{\tilde{\mu}} \\
&\left[\frac{(\beta(1-\alpha)+\alpha)\tilde{\mu}-1}{\beta} L_A^{-1} + \frac{B_{Z,A}(s)}{1+B_{Z,A}(s)} \frac{\partial \ln \phi_Z(s, L_A)}{\partial L_A} + \tilde{\mu} \frac{B_{Z,A}(\tilde{s})}{1+B_{Z,A}(\tilde{s})} \frac{\partial \ln \phi_Z(\tilde{s}, L_A)}{\partial L_A} \right]
\end{aligned}
\tag{A.33}
$$

where $\tilde{C} = \left(\frac{(1-\mu)}{\overline{H}}\right)^{\mu-1} \left(A^{\frac{1}{1-\alpha}+\frac{\beta-1}{\beta(1-\alpha)+\alpha}} \left[\frac{\alpha}{p_X}\right]^{\frac{\alpha}{1-\alpha}+\frac{\alpha(\beta-1)}{\beta(1-\alpha)+\alpha}} (1-\alpha) \right)^{\mu}$

For a worker with average wage, this expression is negative, provided that:

$$
\frac{(\beta(1-\alpha)+\alpha)\tilde{\mu}-1}{\beta} L_A^{-1} + (1+\tilde{\mu}) \frac{B_{Z,A}(\tilde{s})}{1+B_{Z,A}(\tilde{s})} \frac{\partial \ln \phi_Z(\tilde{s}, L_A)}{\partial L_A} < 0 \tag{A.34}
$$

$\left|\frac{(\beta(1-\alpha)+\alpha)\tilde{\mu}-1}{\beta}\right| > 1+\tilde{\mu} \leftrightarrow -\frac{(\beta(2-\alpha)+\alpha)\tilde{\mu}-1}{\beta} > 1$ $\tilde{\mu} > \frac{\beta-1}{\beta(2-\alpha)+\alpha}$ is satisfied as $-\tilde{\mu} < 0$ by the no black-hole condition (so that I can handle absolute value as I do):

$$
\leftrightarrow -\mu \frac{(1-\alpha)(2\beta-1)+\alpha}{(1-\alpha)(\beta-1)} > \frac{\beta-1}{\beta(2-\alpha)+\alpha} - 1
$$

$$
\leftrightarrow \mu < \frac{1+\beta(1-\alpha)+\alpha}{\beta(2-\alpha)+\alpha} \frac{(1-\alpha)(\beta-1)}{(1-\alpha)(2\beta-1)+\alpha}
$$

which is smaller than the no black-hole condition as $\beta > 1$ (compare with equation (A.26)). The condition is thus more restrictive and tends to the no black-hole condition as $\beta \to 1$.

Together with $\frac{B_{Z,A}(s)}{1+B_{Z,A}(s)} \frac{\partial \ln \phi}{\partial L_A} < \frac{1}{L_A}$ it ensures that equation (A.34) is negative. For workers below average, $\frac{\partial V_A(s)}{\partial L_A}$ is even more negative. For workers above average, this derivative might be positive.[3] For this group of workers, their density is necessarily decreasing in city size and will exactly balanced the above average increase in the term $B_{Z,A}(s)$. Thus if an equilibrium exists with many populated sites, population must be increasing in the exogenous productivity of the site.

Proposition 1 part (ii)

[3]If some high-skilled workers have $\frac{\partial V_A}{\partial L_A} > 0$, note that this would be for small L_A because $\lim_{L_A \to L} \frac{\partial \phi}{\partial L_A} \to 0$

The skill distribution of larger cities first order stochastically dominates the one of smaller cities: $\frac{\partial^2 \frac{L_E(s,A)}{L_E(A)}}{\partial A \partial s} > 0$ and this effect is stronger when the ratio of prices of simple to complex capital $\frac{p_X}{p_Z}$ increases $\frac{\partial^3 \frac{L_E(s,A)}{L_E(A)}}{\partial A \partial s \partial \frac{p_X}{p_Z}} > 0$:

Constant utility across cities (spatial indifferent condition) imply that two cities A and B provide the same utility for a worker with a given skill:

$$r_A^{\mu-1} w_A(s) = r_B^{\mu-1} w_B(s) \tag{A.35}$$

Taking the ratio of equation (A.35) for two skill group s and t imply that $\frac{w_A(s)}{w_A(t)} = \frac{w_B(s)}{w_B(t)}$ or equivalently that $\frac{w_A(s)}{w_A(t)} \frac{w_B(t)}{w_B(s)} = 1$. The increase in wage across cities is thus the same for all skill types and reflects the common increase in rents.

Dividing equation (A.15) for two skill group s and t and two cities A and B gives:

$$\frac{L_A(s)L_B(t)}{L_A(t)L_B(s)} = \left[\frac{(1+B_{Z,A}(s))(1+B_{Z,B}(t))}{(1+B_{Z,A}(t))(1+B_{Z,B}(s))}\right]^{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta} > 1 \tag{A.36}$$

where $B_{Z,A}(s) = \phi_Z(s, L_A)^{\frac{\beta}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}} \left[\frac{p_X}{p_Z}\right]^{\frac{\alpha(\beta-1)}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}}$.

Equation (A.36) is bigger than one since $(1 + B_{Z,A}(s))$ is log supermodular in s and L (Lemma 2 part (i)).

Equation (A.36) is increasing in $\frac{p_X}{p_Z}$. This directly follows from Lemma 2 part (iii).

Proposition 1 part (iii) and (iv)

The firm allocates more computerized equipment in large cities compare to smaller ones to workers with a given skill: $\frac{\partial k_{Z,E}(s,A)}{\partial A} > 0$:

Consider equation (A.17) . The total derivative of the assignment computer capital with the exogenous productivity of the site is:

$$dk_{Z,A}(s) = \frac{\partial k_{Z,A}(s)}{\partial A} dA + \frac{\partial k_{Z,A}(s)}{\partial Y_A} dY_A + \frac{\partial k_{Z,A}(s)}{\partial L_A(s)} dL_A(s) + \frac{\partial k_{Z,A}(s)}{\partial \phi_Z(s, L_A)} \frac{\partial \phi_Z(s, L_A)}{\partial L_A} dL_A > 0 \tag{A.37}$$

By inspection, the first two partial derivatives are positive: $\frac{\partial k_{Z,A}(s)}{\partial A} > 0$, $\frac{\partial k_{Z,A}(s)}{\partial Y_A} > 0$. As city size increases with the exogenous productivity of the site A, as well of the share of high skilled workers (Proposition 1 part (i) and (ii) , those terms go in the sense of having more computer capital assigned for a given worker the larger the cities. By contrast, for workers with increasing $dL_A(s)$ in city size, $\frac{\partial k_{Z,A}(s)}{\partial L_A(s)} < 0$ leads to a *decrease* in computer capital as city size increases. Nevertheless, by the spatial equilibrium condition this term cannot dominate. Indeed, would that term dominate for those workers, their wage would be decreasing in city size as $L_A(s)$ enters the wage and capital at the same rate $\frac{-1}{\beta(1-\alpha)+\alpha}$. This is not possible since housing costs increase in city size, every worker must be compensated by a higher wage there to accept living there.

Finally, I calculate $\frac{\partial \log k_{Z,A}(s)}{\partial \phi_Z(s,L_A)}$, which is positive (then, a fortiori will be $\frac{\partial k_{Z,A}(s)}{\partial \phi_Z(s,L_A)}$):

$$\frac{\partial \log k_{Z,A}(s)}{\partial \phi_Z(s,L_A)} = \frac{\beta \phi_Z(s,L_A)^{-1}}{(1-\delta)\left[\beta(1-\alpha)+\alpha\right]+\delta}\left[1-\frac{\delta(1-\alpha)(\beta-1)}{\beta(1-\alpha)+\alpha}\frac{B_{Z,A}(s)}{(1+B_{Z,A}(s))}\right] > 0 \tag{A.38}$$

Consequently, the whole derivative is positive.

Regarding the derivative of simple capital per worker $k_{X,A}(s)$ with respect to city size L_A, the effect is theoretically ambiguous:

$$dk_{X,A}(s) = \frac{\partial k_{X,A}(s)}{\partial A}dA + \frac{\partial k_{X,A}(s)}{\partial Y_A}dY_A + \frac{\partial k_{X,A}(s)}{\partial L_A(s)}dL_A(s) + \frac{\partial k_{X,A}(s)}{\partial \phi_Z(s,L_A)}\frac{\partial \phi_Z(s,L_A)}{\partial L_A}dL_A ? 0 \tag{A.39}$$

The only difference between simple and computer capital is the last derivative $\frac{\partial k_{X,A}(s)}{\partial \phi_Z(s,L_A)} < 0$. Two effects go thus in opposite direction here. On the one hand, more productive sites and more local GDP lead to more simple capital. On the other hand, the complementarity with simple tasks leads to less simple capital as city size increases.

Note: elasticity of computer capital's allocation with city size as a function of the skill level

The percentage increase in computer capital's allocation with city size might intensify or weaken with the skill of the worker: $\frac{\partial^2 \log k_{Z,A}(s)}{\partial s \partial L_A} \gtrless 0$. If $\frac{\partial^2 \log k_Z}{\partial s \partial A} > 0$, it points towards a stronger impact of computerization on skill sorting than when $\frac{\partial^2 \log k_Z}{\partial s \partial A} \leq 0$.

The allocation of the logarithm of simple capital per worker is always decreasing in s and L $\frac{\partial^2 \log k_{X,A}(s)}{\partial s \partial L_A} < 0$.

Take the ratio of equation (A.17), for two skill groups $s > t$ and two cities $A > B$:

$$\frac{k_{Z,A}(s)k_{Z,B}(t)}{k_{Z,A}(t)k_{Z,B}(s)} = \left[\frac{\phi_Z(s,L_A)\phi_Z(t,L_B)}{\phi_Z(t,L_A)\phi_Z(s,L_B)}\right]^{\frac{\beta}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}} \left[\frac{(1+B_{Z,A}(s))(1+B_{Z,B}(t))}{(1+B_{Z,A}(t))(1+B_{Z,B}(s))}\right]^{\frac{-\delta(1-\alpha)(\beta-1)}{\beta(1-\alpha)+\alpha}} \left[\frac{L_A(s)L_B(t)}{L_A(t)L_B(s)}\right]^{\frac{-1}{\beta(1-\alpha)+\alpha}} \tag{A.40}$$

Plugging the spatial equilibrium condition A.36 in equation (A.40) gives:

$$\frac{k_{Z,A}(s)k_{Z,B}(t)}{k_{Z,A}(t)k_{Z,B}(s)} = \left[\frac{\phi_Z(s,L_A)\phi_Z(t,L_B)}{\phi_Z(t,L_A)\phi_Z(s,L_B)}\right]^{\frac{\beta}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}} \left[\frac{(1+B_{Z,A}(s))(1+B_{Z,B}(t))}{(1+B_{Z,A}(t))(1+B_{Z,B}(s))}\right]^{-1} \tag{A.41}$$

The percentage change in computerized equipment across city size and skill group is equal to the percentage change in time allocated to complex tasks across city size and sill group. The link between the 2 first comes from the optimality condition for computer capital (equation (A.17)), where the optimal amount of computer capital positively depend on the share of time allocated to complex tasks to the power of delta. The relationship between the 2 is though not proportional, where computer capital varies more in response to a given variation in the share of tasks allocated to complex tasks. Indeed, computer capital is not upper bounded as it the share of time allocated to complex tasks. The second link comes from the spatial equilibrium condition. The allocation of computer capital to workers in a skill group also *negatively* depends on the local share of workers in that group. This reduces the percentage increases in computerized equipment across city size with skill, up to the point where it follows the pattern of share of time allocated to complex tasks.

$$\frac{k_{Z,A}(s)k_{Z,B}(t)}{k_{Z,A}(t)k_{Z,B}(s)} = \frac{l_{Z,A}(s)l_{Z,B}(t)}{l_{Z,A}(t)l_{Z,B}(s)} >?1 \tag{A.42}$$

where $l_{Z,A}(s) = \dfrac{B_{Z,A}(s)}{1+B_{Z,A}(s)}$ and $B_{Z,A}(s) = \phi_Z(s,L_A)^{\frac{\beta}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}} \left[\frac{p_X}{p_Z}\right]^{\frac{\alpha(\beta-1)}{(1-\delta)[\beta(1-\alpha)+\alpha]+\delta}}$

The amount of computer capital per worker thus inherits the properties of the share of time allocated to complex tasks (Lemma 1).

By symmetry for simple capital, take the ratio of equation (A.18), for two skill groups $s > t$ and two cities $A > B$:

$$\frac{k_{X,A}(s)k_{X,B}(t)}{k_{X,A}(t)k_{X,B}(s)} = \frac{(1+B_{Z,A}(s))(1+B_{Z,B}(t))}{(1+B_{Z,A}(t))(1+B_{Z,B}(s))} \tag{A.43}$$

$$\frac{k_{X,A}(s)k_{X,B}(t)}{k_{X,A}(t)k_{X,B}(s)} = \frac{(1-l_{Z,A}(s))(1-l_{Z,B}(t))}{(1-l_{Z,A}(t))(1-l_{Z,B}(s))} < 1 \tag{A.44}$$

The amount of simple capital allocated to a worker thus follows the pattern of the percentage share of time allocated to simple tasks.

A.2 Construction and representativeness of the samples

I construct my German sample as follow. Beginning with 20'000 observations I drop the ones with inconsistent occupation "number" (91 observations). Moreover, from the remaining observations, missing values for the industry amounts to 159 observations, missing values for the education to 6 observations, for age 38, and for firm size 656 observations as well as for variables capturing complex task (267 missing values)[4]. I drop observations pertaining to the agricultural sector due to the fact that very few live in large cities and that their main factor of production is land, which I abstract from in the model (316 observations). Nevertheless, as it accounts for only 5% of the sample it does not affect the results. Overall, the final sample includes 18'467 workers.

I construct my sample for Great Britain as follow. Beginning with $7,787$, I exclude Northern Ireland because it is not available in larger dataset I use to compare the consistency of the BSS (498 observations). Then, missing observations for occupation and industries, as well as armed force and agriculture and fishery amount to 119 observations. Last, missing observations for education, firm size categories and the complexity of tasks amount to 10, 39, 111 respectively, which leaves me with $7,010$ observations.

Table A.1 compares the number of observations for 7 population categories of agglomerations (Travel to work areas), where small agglomerations are slightly oversampled.

A.3 Share of high skilled workers in large cities

This appendix replicates with my datasets the main stylized fact in which this paper originated: the sorting of high skilled workers in large cities.

Figure A.1 below shows the density of workers with and without a job requiring a university degree across agglomeration density in the German BIBB dataset. Figure A.2 displays the four education groups according to the requirement of the occupation. Larger

[4] percentage time using computer as well as an interactive and analytical tasks index, see Section 1.4.1

Table A.1: Observations on 6 TTWA categories in Great Britain

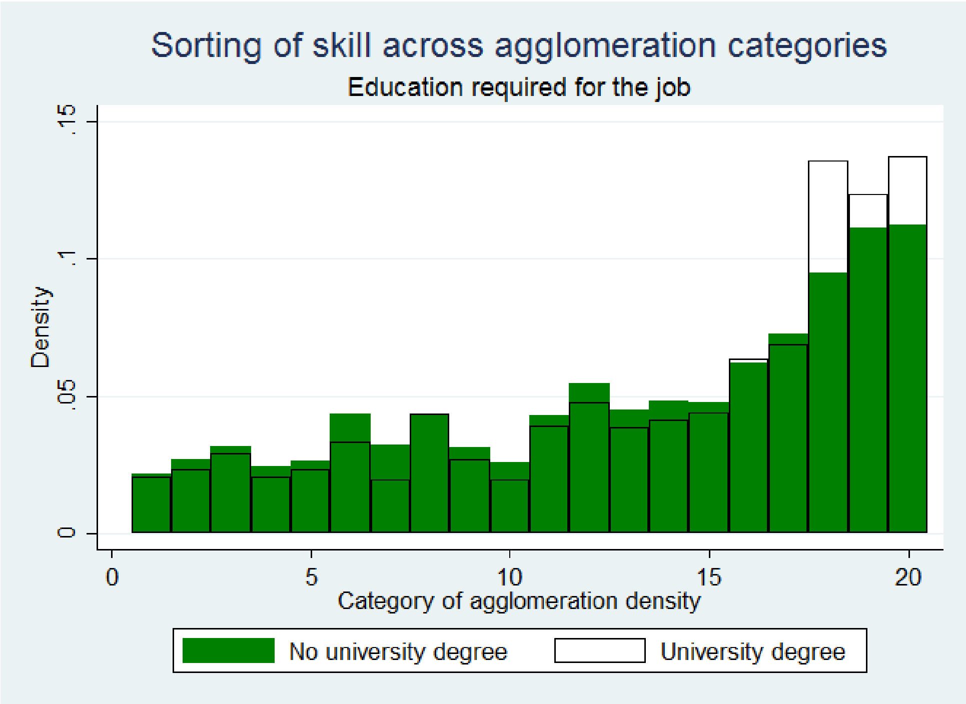
Log of population 1991 census	Working residents BSS	Percent	Residents 1991 census	Percent
≤ 10	309	4.4%	$780'277$	1.5%
(10, 11]	638	9.1%	2,895,963	5.5%
(11, 12]	1,548	22.1%	8,804,636	16.8%
(12, 13]	1,818	25.9%	14,132,168	27%
(13, 14]	1,736	24.8%	13,051,460	24.9%
(14, 15]	612	8.7%	6,154,277	11.8%
> 15	349	5%	6,521,442	12.5%
Total	7,010	100%	52,340,223	100%

Source: author's computation based on the 2006th wave of the British Sill Survey (7,010 observations). *Notes:* rural areas, which I define as TTWAs with less than $60'000$ inhabitants are the first 2 categories. Large cities are defined as TTWAs with more than $430'000$ inhabitants (categories 5, 6, 7). The last category only includes London.

cities witness a higher proportion of high-skilled jobs.[5]

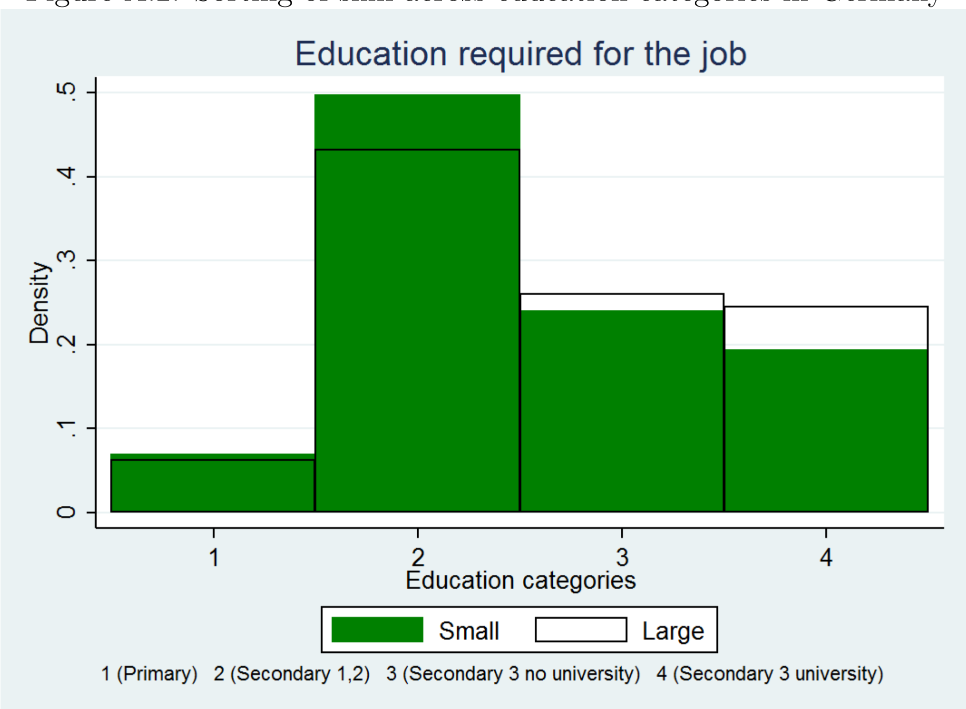
[5]The first group corresponds to primary education, the second group comprises the first and second stages of secondary education or apprenticeship, the third group represents about 4 year of education after the age of 17 or 18 not equivalent to a university degree, the last one is a job requiring a university degree.

Figure A.1: Sorting of skill across agglomeration categories in Germany



Source: author's computation based on the 2006th wave of the German working population survey.

Figure A.2: Sorting of skill across education categories in Germany



Source: author's computation based on the 2006th wave of the German working population survey.

Next, Table A.2 compares the share workers working in an occupation requiring a university degree in the British Skill Survey (7′010 observations) and the larger Annual

Survey of Hours and Earnings (200′000 observations) across 3 groups of Travel to Work Areas: Urban centers, small towns and rural areas. The whole sample of the British Skill Survey slightly underestimate the proportion of workers with a university degree, whereas the subsample used for the wage analysis slightly overestimate it in all settlement types. See Appendix A.12 for a description of the Annual Survey of Hours and Earnings.

Table A.2: Workers in an occupation requiring a university degree in Great Britain

(a) BBS: Whole sample 7,010 observations

	Whole sample	Large cities	Small towns	Rural areas
Frequency	24%	27%	23%	19%
Observations	*7,010*	*2,697*	*3,366*	*947*

(b) BBS: wage subsample 4,117 observations

	Whole sample	Large cities	Small towns	Rural areas
Frequency	27%	30%	26%	22%
Observations	*4,117*	*1,641*	*1,960*	*516*

(c) ASHE

	Whole sample	Large cities	Small towns	Rural areas
Frequency	26%	28%	25%	20%
Observations	*243,043*	*128,260*	*102,131*	*12,652*

Source: author's computation based on the 2006th wave of the British Sill Survey as well as the 2006th wave of the Annual Survey of Hours and Earnings.
Notes: rural areas are TTWAs with less than 60′000 inhabitants. Large cities are TTWAs with more than 440′000 inhabitants.

A.4 Elasticity of computer capital's allocation with city size and skill

The behavior of the elasticity of computer capital's allocation with skill and city size cannot be a distinguishing statistics for the main story of this paper in general. Nevertheless, provided that the percentage increase in computer capital's allocation with city size intensifies with the skill of the worker, one can conclude that computerization is one explanation for the sorting of high-skilled workers in large cities. This condition is though not necessary. Indeed, as the optimal amount of computer capital depends on the optimal share of time allocated to complex tasks, the decreasing returns to tasks also impacts computer capital allocation. Because high-skilled workers spent overall a larger share of time to complex tasks, they are more impacted by this effect. Moreover, perfect mobility together with local skill complementarity also goes in the direction of allocating less computer capital with city size as the skill of the worker increases. These effects dominate if $\frac{\partial^2 \ln \phi_Z(s,L_A)}{\partial L_A \partial s} = 0$ or if the log supermodularity of $\phi_Z(s, L_A)$ in s and L_A is relatively small (see Assumptions 1).

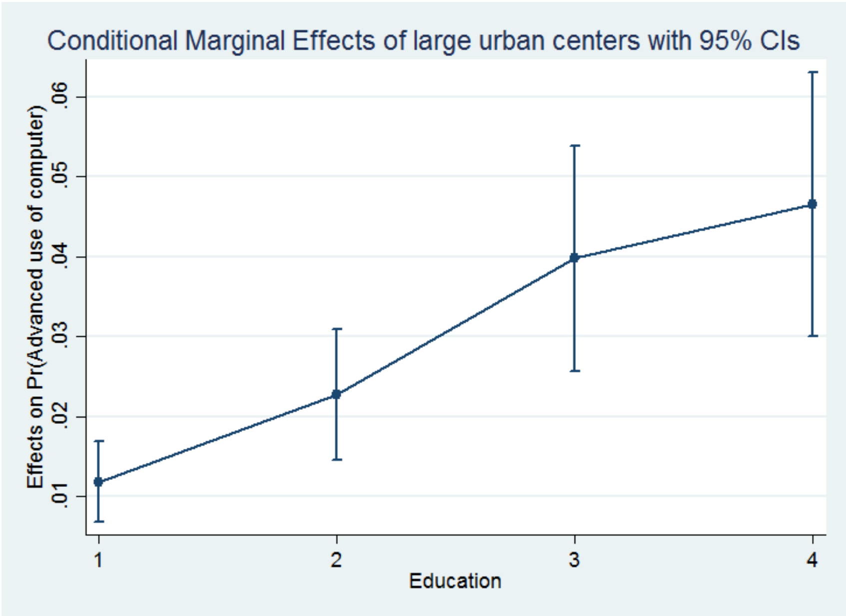
In order to document the empirical direction of part (iv) of Proposition 1, I evaluate equation (1.12) at different skill level. I face the same challenges as with the test of part (iii) of Proposition 1: I use the same proxy for computer capital and I also evaluate the version of equation (1.11) with city size instrumented. Last, I measure skill with the first digit of the occupation category.[6] Unobserved skill is a threat to the conclusion that significantly different coefficients imply a heterogeneous correlation of computer capital and city size given the skill of workers. The condition for an interpretation "given the skill of workers" is though weaker than above. Workers can sort across city sizes according to unobserved characteristics, but the sorting must be the same across observed categories as I compare them. Based on UK and German data, I find a decreasing percentage increase with the education of the worker in the probability to have an advanced use of computer at work.

Figures A.3 shows the coefficient γ_1 of equation (1.12) evaluated at different skill level. Since it is a marginal effect, it is in percentage points. The pattern is weaker when formal education is used instead of education required for the job from the first digit of the occupation category. Dividing those coefficients by the conditional probability to have an advanced use of computer in non large urban areas gives the percentage increase in the coefficient of interest with observed skill. It is decreasing in skill, in all specification.

Alternatively, I show the pattern of the percentage working time on computer since the pattern of computer capital in part (iv) of Proposition 1 follows the pattern of complex tasks.

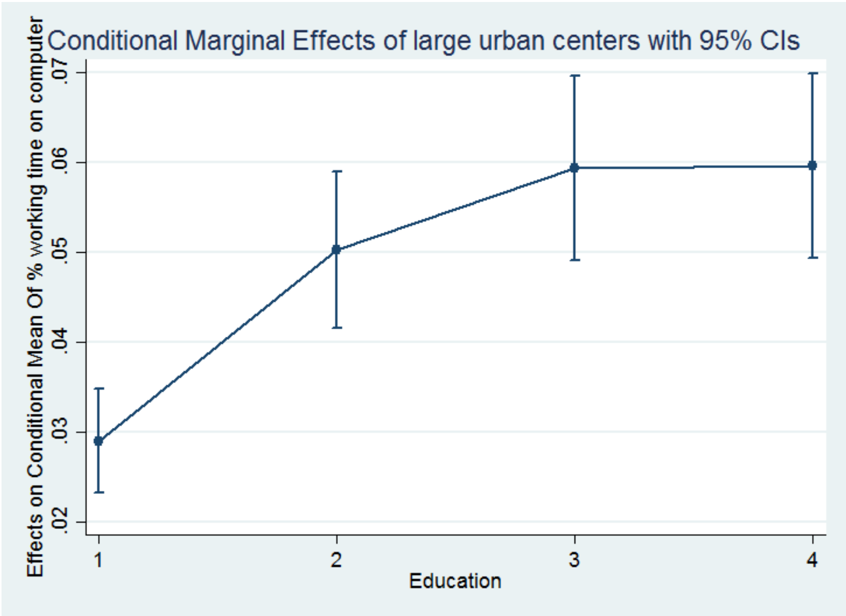
[6](1) primary education (ISCED 1), (2) ISCED 2 and 3, (3) ISCED 5, ISCED 6 and 7 (university degree) (4)

Figure A.3: City and advanced use of computer in Germany: education



Source: author's computation based on the 2006th wave of the German working population survey
Notes: Education levels are based on the first digit of occupation.
Dividing the coefficient by the predicted probability in small towns and rural areas gives the coefficient in percentage, which is decreasing in skill: (1)2.1% (2) 1.9%(3) 1.6%(4)1.4%

Figure A.4: City and percentage working time on computer in Germany: education



Source: author's computation based on the 2006th wave of the German working population survey
Notes: Education levels are based on the first digit of occupation.
Dividing the coefficient by the predicted probability in small towns and rural areas gives the coefficient in percentage, which is decreasing in skill: (1)2.3% (2) 1.8%(3) 1.3%(4)1.3%

A.5 Construction of my instrument: long lagged population

I use data from the 1851 census available at the UK data archives to construct my instrument (Southall et al. (2004)). I match historic places with geo-coordinates using the Ordnance Survey Names database. I then use Geographical Information System (GIS) to assign historic places to travel to work areas according to commuting flows from the 1991 census. I assign historic places without a match in the OSNames database to the same current travel to work area as historic places in the same historic subdistrict or district.

A.6 Advanced computer and city size: generalized ordered logit

This appendix uses the 4 possible values available in the survey for computer use of worker i c_i instead of creating a dummy. Because the values of the variables can be ordered (1.Not working with a computer; 2. Working with a computer without writing programs or using macro; 3. Workers writing programs or using macro, and only as users; 4. Workers using computers beyond user activities.), I use a generalized ordered logistic model.[7] As the

[7] See Williams (2006).

parallel line assumption is not rejected for the large city dummy, I constraint its coefficient:

$$P(c_i > j|x_i) = G(\gamma_{0,j} + \gamma_1 Pop_j + \gamma_2 Ind_i), \quad j = 1, 2, 3 \tag{A.45}$$

where the variables are the same than in equation (1.12) and G is the logistic cumulative distribution function.

Table A.3: Computer use and city size in Germany (generalized ordered logistic marginal effects at mean workers' characteristics)

Indep. var.	Dep. var.[a/]	(1) Whole sample	(2) No rural	(3) No large cities	(4) No university[b/]	(5) University[b/]
Large city	1	-.029***	-.025***		-.052***	-.003**
		(.0033)	(.0035)		(.0060)	(.0015)
	2	-.016***	-.017***		.007***	-.025**
		(.0020)	(.0026)		(.0013)	(.0113)
	3	.020***	.018***		.021***	.011**
		(.0023)	(.0027)		(.0024)	(.0048)
	4	.025***	.023***		.024***	.018**
		(.0028)	(.0033)		(.0028)	(.0080)
Small towns vs rural	1			-.009		
				(.0054)		
	2			-.002		
				(.0012)		
	3			.005		
				(.0031)		
	4			.006		
				(.0035)		
Worker's characteristics[c/]		Yes	Yes	Yes	Yes	Yes
Occupation and industry		No	No	No	No	No
Observations		18,467	14,773	9,546	13,137	5,330

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

[a/] The dependent variable takes the following values: 1.Not working with a computer; 2. Working with a computer without writing programs or using macro; 3. Workers writing programs or using macro, and only as users; 4. Workers using computers beyond user activities.

[b/] According to formal education.

[c/] Age, sex and 4 education dummies including primary education, secondary 1 and 2, secondary 3 without a university degree and secondary 3 with a university degree.

Source: author's computation based on the 2006th wave of the German working population survey.

Notes: Large cities, small towns and rural areas are defined according to the German administration http://www.bbsr.bund.de/BBSR/DE/Raumbeobachtung/Raumabgrenzungen/StadtLandRegionen_Typen/StadtLandRegionen_Typen.html?nn=443270, visited the 14th of January 2017.

The odds ratios are found to be constant for all values of the dependent variable and equal to (robust standard errors in parenthesis): column 1: 1.3 (.038), column 2: 1.27 (.043), column 3: 1.07 (.045, pvalue=.1), column 4 : 1.36 (.047), column 5:, 1.13 (.064).

Table A.3 confirms the results of the benchmark case in Table 1.2. The estimates in column (1) imply that the probability of a worker with mean characteristics not to use a computer at work is 11.2% in urban centers and 14.1% in the rest of the country. The probability to use a computer without writing programs or using macro is 64.5% in urban centers and 66.1% in the rest of the country. By contrast, working with programs or using macro: 12.4% in urban centers versus 10.4% in the rest of the country. Last, workers programming for others are 11.9% percent in large urban centers compared to 9.4% in the rest of the country. The qualitative results holds for both workers with and without a university degree, and for the sample without rural areas. As in Table 1.2, their is no significant differences in computer capital between small town and rural areas.

A.7 Categories of city density in Germany

The categories in Table A.4 represents each 5 percentiles of the distribution with the first category describing the less dense area.

Table A.4: Categories of city density in Germany

Number of category	Density range (inhabitant / km^2)
1	0-71
2	72-97
3	98-109
4	110-115
5	116-120
6	121-139
7	140-153
8	154-161
9	162-174
10	175-182
11	183-196
12	197-213
13	214-239
14	240-256
15	257-273
16	274-370
17	371-456
18	457-715
19	716-1080
20	above 1080

Note: The categories represent each fifth percentile of the distribution of city density.

A.8 Further robustness

A.8.1 Alternative proxies for computer capital

Data and overview

This appendix presents 2 additional proxies for computer capital in equation (1.12): "having a computer at work", and "working with a computer as main equipment". Having a computer at work is a straightforward proxy for computer capital. Nevertheless, as 98% of the workers employed in an occupation requiring a university degree regardless of location have a computer at work, it offers few variation of computer capital for this labor group. I construct my second alternative proxy as a dummy variable indicating among 14 work equipments which one is the most important. This variable in the survey is constructed based on an open question. It offers more variation for workers with a university degree, where 61% have computer as their main work equipment.

Table A.5 shows that the means of these two additional proxies for computer capital are also higher in large urban centers relative to the rest of the country. Moreover, the prevalence of computer at work is also higher in Germany relative to Great Britain.

Working with a computer

Germany

Table A.6 shows the coefficient of interest γ_1 from equation (1.12) with working with a computer as a proxy for computer capital. In the benchmark, reported in column (1), the probability to work with a computer goes from 89.1% in large urban centers to 85.6% in the rest of the country for a worker with mean characteristics. Columns (2) and (3) show that most of the effect takes place between small agglomerations and large urban centers. Nevertheless, contrary to Table 1.2 the coefficient in column (3) is slightly significant. Column (4) and (5) indicate that the effect is almost entirely driven by workers without a university diploma. As 98% of workers with a university degree do work with a computer, I probably lack the variation necessary for an estimation for this group.[8] Finally, column (6) takes the perspective of the job, controlling for occupations dummies at 3 digits level and for industries at 2 digits level. The results are thus coherent with Section 1.3.3 with 2 exceptions. First of all, there is no significant relationship between computer and large urban centers when restricting the sample to workers with a university degree. Second, there exists a slightly significant relationship (at the 10% level) between small towns and the probability to have a computer at work.

[8]The stronger effect for low-skilled might either be that the proxy having a computer capital at work is not capturing change in computer capital allocation for high-skilled, or that there is no effect for them.

Table A.5: Measure of computer capital: descriptive statistics

(a) Having a computer at work (dummy): Germany

Mean (StD) *Observations*	Whole sample (1)	Large cities[b] (2)	Rest of the country (3)	$\|t$-stat$\|$ [c] (4)
Whole sample	.83	.86	.75	9.3***
	(.38)	(.35)	(.43)	
	18,467	*8,921*	*9,546*	
University degree[a]	.98	.98	.97	2**
	(.15)	(.14)	(.16)	
	4,951	*2,675*	*2,276*	
No university degree[a]	.77	.80	.75	7***
	(.42)	(.40)	(.43)	
	13,516	*6,246*	*7,270*	

(b) Having a computer at work (dummy): Great Britain

Mean (StD) *Observations*	Whole sample (1)	Large cities[c] (2)	Rest of the country (3)	$\|t$-stat$\|$ [c] (4)
Whole sample	.73	.75	.71	4.2***
	(.45)	(.43)	(.45)	
	7,010	*2,697*	*4,313*	
University degree[a]	.95	.96	.94	1.9*
	(.22)	(.20)	(.24)	
	1,689	*725*	*964*	
No university degree[a]	.65	.68	.64	2.7***
	(.48)	(.47)	(.50)	
	5,321	*1,972*	*3,349*	

(c) Computer main equipment at work (dummy): Germany

Mean (StD)	Whole sample (1)	Large cities[b] (2)	Rest of the country (3)	$\|t$-stat$\|$ [c] (4)
Whole sample	.45	.49	.41	11.5***
	(.50)	(.50)	(.49)	
University degree[a]	.61	.64	.57	5.3***
	(.49)	(.48)	(.50)	
No university degree[a]	.39	.43	.36	8.3***
	(.49)	(.50)	(.48)	

[a] Education required for occupation.
[b] As defined by to the German administration http://www.bbsr.bund.de/BBSR/DE/Raumbeobachtung/Raumabgrenzungen/StadtLandRegionen _Typen/StadtLandRegionen_Typen.html?nn=443270, visited the 14th of January 2017.
[c] T-statistic in absolute value testing equalities of means in column (2) and (3), assuming unequal variance, with *** p<0.01, ** p<0.05, * p<0.1.
[d] TTWAs with more than 440′000 inhabitants.
Source: author's computation based on the 2006th waves of the British Sill Survey and the German working population survey

Table A.6: Computer use and city size in Germany (logistic marginal effects at mean workers' characteristics)

Dep. var. Pr(computer) VARIABLES	(1) Whole sample	(2) No rural	(3) No large cities	(4) No university[a/]	(5) University[a/]	(6) Whole sample[b/]
Large city	.035***	.029***		.048***	.005*	.005*
	(.0046)	(.0049)		(.0070)	(.0025)	(.0033)
Small towns vs rural			.013*			
			(.0071)			
Worker's characteristics[c/]	Yes	Yes	Yes	Yes	Yes	No
Occupation and industry fixed effects	No	No	No	No	No	Yes
Observations	18,467	14,773	9,546	13,516	4,951	17,186

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
[a/] Education required for occupation.
[b/] Some observations drop because predicting success or failure perfectly. Column (6) controls for 105 occupations (3-digits) and 61 industries (2-digits).
[c/] 4 education dummies including primary education, secondary 1 and 2, secondary 3 without a university degree and secondary 3 with a university degree, age and sex.
Source: author's computation based on the 2006th wave of the German working population survey.
Notes: Large cities, small towns and rural areas are defined according to the German administration.
http://www.bbsr.bund.de/BBSR/DE/Raumbeobachtung/Raumabgrenzungen/StadtLandRegionen_Typen/StadtLandRegionen_Typen.html?nn=443270, visited the 14th of January 2017.

Great Britain

Table A.7 shows the coefficient of interest γ_1 from equation (1.12) with working with a computer as a proxy for computer capital. I include column (1) for comparison with the German estimates. It shows that a worker with mean characteristics has a probability of 78.6% instead of 75.1% to work with a computer in large cities relative to the rest of the country. The estimate of column (2) implies that a worker with mean characteristics in a city with mean size has a probability of 78.5% instead of 76.4% of having an advanced use of computer at work when city size increases by 100%. Columns (3) and (4) show that most of the effect takes place between small agglomerations and large urban centers. Column (5) and (6) indicate that the effect is entirely driven by workers without a university diploma. Finally, column (7) takes the perspective of the job, controlling for occupation and industry dummies. Here, the coefficient is insignificant (though I control for 4 digit occupations, contrary to table A.6, which only controls for 3 digits occupations). Comparing with advanced computer use as a dependent variable, the coefficient for computer use is slightly lower for the German case and larger for the British case. The magnitude of the coefficient between the German and British datasets looks similar.

Table A.7: Computer use and city size in Great Britain (logistic marginal effects at mean workers' characteristics)

Dep. var. Pr(computer) VARIABLES	(1) Whole sample	(2) Whole sample	(3) No rural	(4) No large cities	(5) No university[a/]	(6) University[a/]	(7) Whole sample[b/]
Large city	.033***						
	(.0137)						
ln pop		.0205***	.0138***	.0370***	.0217***	.0021	.0042
		(.0039)	(.0054)	(.0070)	(.0052)	(.0026)	(.0054)
Worker's characteristics[c/]	Yes	Yes	Yes	Yes	Yes	Yes	No
Occupation and industry fixed effects	No	No	No	No	No	No	Yes
Observations	7,010	7,010	6,063	4,313	5,321	1,689	5,735

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

[a/] Education required for occupation.

[b/] Some observations drop because predicting success or failure perfectly. Column (6) controls for 338 occupations (4-digits) and 9 industries (1-digits).

[c/] Education (5 categories according to the National Vocational Qualification levels), age and sex.

Source: author's computation based on the 2006th wave of the British Sill Survey.

Notes: Ln pop is the log of population by travel to work area (TTWA) which represents an integrated labour market based on commuting flows. Rural areas are defined as TTWA with less than $60,000$ inhabitants and large TTWAs are defined as TTWAs with more than $440,000$ inhabitants.

Computer as the main equipment (Germany)

Table A.8 shows the coefficient of interest γ_1 from equation (1.12) with working with a computer a main equipment as a proxy for computer capital. In the benchmark, reported in column (1), the probability to have computer as ones main equipment goes from 48.8% in large urban centers to 41.4% in the rest of the country for a worker with mean characteristics. The estimate almost doubles compared to having advanced computer use as dependent variable (Table 1.2). Columns 2 and 3 show that most of the effect takes place between small agglomerations and large urban centers, even though the effect is also statistically significant between small towns and rural areas. The effect for workers in occupations requiring a university or not is similar in magnitude (column (4) and (5) shows a stronger effect for workers with a university degree. Finally, column (7) shows that there is a within occupation variation in the probability of having computer as ones main work equipment. That is, the same job is done differently in large urban centers relative to the rest of the country.

Table A.8: Computer as main equipment and city size in Germany (logistic marginal effects at mean workers' characteristics)

Dep.var.	(1)	(2)	(3)	(4)	(5)	(6)
Pr(computer main equipment)						
VARIABLES	Whole sample	No rural	No large cities	No university[a/]	University[a/]	Whole sample[b/]
Large city	.074***	.064***		.063***	.068***	.020**
	(.0075)	(.0086)		(.0087)	(.0142)	(.0098)
Small towns vs rural			.025**			
			(.0106)			
Worker's characteristics[c/]	Yes	Yes	Yes	Yes	Yes	No
Occupation and industry fixed effects	No	No	No	No	No	Yes
Observations	18,467	14,773	9,546	13,516	4,951	18,184

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
[a/] Education required for occupation.
[b/] Some observations drop because predicting success or failure perfectly. Column (6) controls for 105 occupations (3-digits) and 61 industries (2-digits). With 4 digits - 298 occupations, the city coefficient of .015 is slightly non significant (p=0.16, 17,451 observations).
[c/] 4 education dummies including primary education, secondary 1 and 2, secondary 3 without a university degree and secondary 3 with a university degree, age and sex.
Source: author's computation based on the 2006th wave of the German working population survey.
Notes: Large cities, small towns and rural areas are defined according to the German administration http://www.bbsr.bund.de/BBSR/DE/Raumbeobachtung/Raumabgrenzungen/StadtLandRegionen_Typen/StadtLandRegionen _Typen.html?nn=443270, visited the 14th of January 2017.

A.8.2 Analysis by industry

Table A.9 describe the results of equation (1.12) by broad industries in the German BIBB dataset. It shows that in the public sector is the only sector with no link between the probability to have an advanced use of computer at work and the large city dummy. Table A.10 shows the same disaggregation by broad industries even though the classification differs. The public sector also shows no effect, as well as distribution and restoration, where those two industries have most of the observations. As in Germany, manufacturing and services display the strongest effect, though none of the coefficient is significant at the 1% level, which might be due to the lower number of observations.

Table A.9: Advanced computer use and city size in Germany (logistic marginal effects at mean workers' characteristics): industries

Dep. var. Pr(advanced computer) VARIABLES	(1) Public sector	(2) Manufacturing	(3) Artisanry	(4) Trade	(5) Other services
Large city	.008	.063***	.027*	.033**	.048***
	(.011)	(.016)	(.015)	(.015)	(.014)
Worker's characteristics[a/]	Yes	Yes	Yes	Yes	Yes
Occupation and industry	No	No	No	No	No
Observations	5,138	3,912	1,901	2,062	4,248
Mean advanced use of computer	.21	.33	.14	.16	.28
% obs in urban centers	48	46	38	48	56

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
[a/] 4 education dummies including primary education, secondary 1 and 2, secondary 3 without a university degree and secondary 3 with a university degree, age and sex.
Source: author's computation based on the 2006th wave of the German working population survey
Notes: Large cities are defined according to the German administration http://www.bbsr.bund.de/ BBSR/DE /Raumbeobachtung/Raumabgrenzungen/StadtLandRegionen_Typen/StadtLandRegionen_Typen.html?nn=443270, visited the 14th of January 2017.

Table A.10: Advanced computer use and city size in Great Britain (logistic marginal effects at mean workers' characteristics): industries

Dep. var. Pr(advanced computer) VARIABLES	(1) Public admin, educ, health	(2) Manufacturing	(3) Construction	(4) Distribution, hotels, restaurant	(5) Transport, communication	(6) Banking, finance, insurance	(7) Other services
Ln pop	.004	.014*	.011	.003	.012	.021**	.014
	(.0059)	(.0085)	(.0075)	(.0041)	(.0096)	(.0101)	(.0092)
Worker's characteristics[a/]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation and industry	No	No	No	No	No	No	No
Observations	2,383	962	391	1,133	476	1,037	329

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

[a/] Education (5 categories according to the National Vocational Qualification levels), age and sex.

Source: author's computation based on the 2006th wave of the British Sill Survey.

Notes: Ln pop is the log of population by travel to work area (TTWA) which represents an integrated labour market based on commuting flows.

A.9 Comparison Germany and Great Britain

This appendix compares the German and British coefficients based on increase in city size and city rank. The coefficient in the German sample can be understood as an increase in density between the agglomeration with median population density in the large cities group (München, 471 inhabitants/km^2) compared to an agglomeration with median population density in the rest of the country (Bremen-Umland, 154 inhabitants/km^2). This entails an increase in about 200% in city density. However, the coefficient of the Great Britain dataset is for an increase in population. I thus also consider the the percentage increase in rank between these two agglomerations: 50%[9] This is also not perfect as the distribution of cities is not completely comparable in the German and Great British case, but it gives an indication. For the case of Great Britain, the mean city at which the coefficient is evaluated has about $270,000$ inhabitants (TTWA: e.g. Swansea). An increase in 200% percent is about $810,000$ inhabitants, e.g. somewhere between Sheffield and Rotherham ($743,010$) and Liverpool ($980,000$ inhabitants). The change in rank is about: 32%[10] The Great British coefficient in Table 1.4 column (1) for an increase in city size of 200% is an increase in the probability to have an advanced use of computer of .028 percentage points. It is a bit lower than the estimate of Table 1.2 which is an increase in .040 percentage points. On the contrary, based on this way of comparison, the increased probability to use a computer at work is larger in the British dataset than the German one (.041 versus .035 percentage points, see Tables A.6 and A.7 column (1) in Appendix A.8.1.)

[9] The rank of München is 86, the one of Bremen-Umland is 36, divided by the total number of agglomerations: 97.

[10] Rank of Sheffield and Rotherham 290, rank of Swansea 196 and number of TTWA: 297. Note that there are many more small agglomerations in Great Britain than the Germany by construction.

A.10 Complex tasks and city size

This appendix shows that there exists a variation in the tasks content of occupation across city sizes given workers' characteristics. Because my tasks variables are indexes, I use a fractional logit response regression[11]:

$$Complex_i = G(\nu_0 + \nu_1 Largecity_j + \nu_2 C_i) \tag{A.46}$$

where $Complex_i$ is the complex tasks index (analytical tasks, interactive tasks and percentage time working with a computer in turn); $G(.)$ is the logistic cumulative distribution function, $Large\ city_j$ is a dummy variable indicating that city j is a city with more than 274 inhabitants per km^2, controls C_i include worker's characteristics - 4 education dummies, age and sex, and later 298 occupations (4-digits) and 61 industries (2-digits).

[11] Papke and Woolridge (1996).

Table A.11: Complex tasks and city size in Germany (fractional logistic marginal effects at mean workers' characteristics)

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. var.	Analytical	Analytical	Interactive	Interactive	% time computer	% time computer
VARIABLES						
Large city	.022***	.008**	.019***	-.002	.063***	.013***
	(.0041)	(.0039)	(.0037)	(.0034)	(.0050)	(.0043)
Worker's characteristics[a/]	Yes	Yes	Yes	Yes	Yes	Yes
Occupation and industry[b/]	No	Yes	No	Yes	No	Yes
Observations	18,467	18,467	18,467	18,467	18,467	18,467

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
[a/] 298 occupations (4-digits) and 61 industries (2-digits).
Source: author's computation based on the 2006th wave of the German working population survey.
Notes: Large cities, small towns and rural areas are defined according to the German administration. http://www.bbsr.bund.de/BBSR/DE/Raumbeobachtung/Raumabgrenzungen/StadtLandRegionen_Typen/ StadtLandRegionen_Typen.html?nn=443270, visited the 14th of January 2017.

Table A.11 shows that all measures of complex tasks are positively related to city size given workers' characteristics (columns (1), (3) and (5)). Next, whereas the relationship seems fully explained by sorting of occupations and industries for interactive tasks (column (4)), it remains when controlling for detailed occupations and industries for analytical tasks and the percentage working time spent on computer (columns (2) and (6)). In the first case, the coefficient is reduced by a factor of 3 and by a factor of 5 in the second case.

A.11 Wage data: Great Britain

This appendix compares the hourly wage data in the 2006th wave of the British Skill Survey with the hourly wage data in the 2006th wave of the Annual Survey of Hours and Earnings across education groups and city sizes. I compare the median of gross hourly wage by groups as well as the share of workers with a university degree across the two datasets. The main differences between the 2 surveys is that the BSS asks the workers their pay, whereas the ASHE is filled in by employers. The quality of the wage data tends to be better in surveys filled in by employers as employee tend to under report high-paid salary and over report the number of worked hours. Nevertheless, over hours not reported by the employee to her employer will miss from the ASHE. Moreover, employers might have an incentive to under report the number of hours for employees at the minimum wage in order to show that they comply with it.

The 2006th wave of the Annual Survey of Hours and Earnings contains 339,861 observations. My main variable of interest is the gross hourly wage, which I construct as dividing the weekly gross wage by the number of weekly working hours including paid overtime. I keep only workers working full time ($243,489$ observations) and drop extreme value for hourly gross wage (value below $1£$ and above $100£$, 190 and 256 observations, respectively). In total I have $243,043$ observations left for my descriptive statistics.

I measure the education of the workers with occupation (standard occupation classification from 2000 - SOC2000), as formal education is absent from the Annual Survey of Hours and Earnings. The education is thus the one required for workers' job. Rural areas are defined as Travel to Work Areas with less than $60,000$ inhabitants according to the 1991 census. Small towns are defined as Travel to Work Areas between $60,000$ and $440,000$ inhabitants. Large cities are Travel to Work Areas with more than $440,000$ inhabitants.

Table A.12 shows that gross hourly wage tends to be smaller in the BSS compared to the ASHE. Also the difference between large cities and rural areas is smaller. This outcome is consistent with top earnings' workers under reporting as they mainly work in large cities. As this under reporting reduces variation across settlement types and skill groups, this should thus make it harder to find any significant effect in Section 1.5. Worrying for me would be if workers with an advanced use of computer in large cities under report their earnings less than other workers with similar level of education.

Next, I compare the city wage premium across the BSS and the ASHE. The smaller

Table A.12: Median wage in Great Britain: descriptive statistics

(a) BBS

Median *Observations*	Whole sample (1)	Urban centers (2)	Small towns (3)	Rural areas (4)
Whole sample	10.0 *4,117*	10.5 *1,641*	9.9 *1,960*	9.0 *516*
University degree	14.8 *1,121*	15.4 *492*	14.8 *516*	12.8 *113*
No university degree	8.8 *2,996*	9.0 *1,149*	8.8 *1,444*	8.1 *403*

(b) ASHE

Median *Observations*	Whole sample (1)	Urban centers (2)	Small towns (3)	Rural areas (4)
Whole sample	10.5 *243,043*	11.3 *128,260*	10.0 *102,131*	8.7 *12,652*
University degree	18.1 *63,525*	19.1 *36,078*	17.1 *24,890*	15.7 *2,557*
No university degree	9.1 *179,518*	9.6 *92,182*	8.9 *77,241*	8.0 *10,095*

Source: author's computation based on the 2006th wave of the British Sill Survey as well as the 2006th wave of the Annual Survey of Hours and Earnings.
Notes: rural areas are TTWAs with less than 60′000 inhabitants. Large cities are TTWAs with more than 440′000 inhabitants. The units of the wage are pounds.

wage difference across city sizes in the BSS compared to the ASHE in Table A.12 translates into a lower urban wage premium in the BSS. The difference between the 2 surveys declines as I add more controls.

Table A.13: Interaction: city size and university on individual wages in Great Britain: BSS vs ASHE

Dep.var. Ln of gross hourly wages$_i$	(1)	(2)	(3)	(4)	(5)	(6)
Indep. var.	BSS	ASHE	BSS	ASHE	BSS	ASHE
Ln pop	.027*	.072***	.019	.057***	.033***	.053***
	(.0154)	(.0128)	(.0118)	(.0103)	(.0089)	(.0107)
Uni			.50***	.65***	.31**	.44***
			(.018)	(.006)	(.126)	(.034)
Uni × Ln pop					.015	.016***
					(.0099)	(.0026)
Observations	4,117	243,043	4,117	243,043	4,117	243,043

Std. Err. adjusted for 178 clusters (TTWA)
*** p<0.01, ** p<0.05, * p<0.1
Source: author's computation based on the 2006th wave of the British Sill Survey as well as the 2006th wave of the Annual Survey of Hours and Earnings.
Notes: The log of population is the log of population by travel to work area which represents an integrated labour market based on commuting flows.

A.12 Total wage bill and individual wages, advanced computer use and city size

In this appendix, I present complementary analysis regarding firms' local demand as measured by their total wage bill across skill group and city size. Indeed, the model implies that allocating computer capital disproportionately increases the productivity of high-skilled workers in large cities. The firms' wage bill is thus more consequent in large cities for high-skilled workers. In the model this is driven exclusively by a larger employment of this group of workers in large cities relative to smaller ones, not by wages. Is that the case that firms have a higher demand for workers in occupations with a more frequent advanced use of computer specifically in large cities? Does the share of workers with a university degree in the occupation add information compared to the share of workers with an advanced use of computer?

For this purpose, I use an alternative data source: the 2006th wave of the Annual Survey of Earnings and Hours. It is one of the largest surveys of earnings in Great Britain (339,861 observations), covering about 1% of the population and is completed by employers which limits measurement errors. Self-employed are excluded from the survey. I measure the labor demand for 2 digits level occupations (SOC2000 classification). Since there are 24 occupations and 297 travel to work areas the number of observations amounts to $7,128$. I construct by occupation and travel to work areas the share of workers with an advanced use of computer as well with a university degree. Table A.14 displays the means by occupation of computer usage, working with a computer and having an advanced use of computer for the British Skill Survey, and the percentage working time on computer and having a computer as main work equipment for the Germann BIBB survey.

Table A.14: Average computer use at work

Occupation SOC2000 2D	Advanced PC	PC at all	% time PC	PC main
11 Corporate managers	.35	.94	43	.60
12 Managers (agriculture & service)	.17	.81	16	.22
21 Science & technology professionals	.72	.96	61	.83
22 Health professionals	.20	.94	30	.23
23 Teaching & research professionals	.25	.94	25	.30
24 Business & public service professionals	.25	.94	53	.70
31 Science & technology associate professionals	.67	.96	58	.64
32 Health & social welfare associate professionals	.10	.87	17	.11
33 Protective service occupations	.16	.96	37	.30
34 Culture, media and sports occupations	.47	.86	49	.58
35 Business & public service associate professionals	.37	.94	57	.66
41 Administrative occupations	.27	.97	63	.70
42 Secretarial & related occupations	.21	.96	70	.87
51 Skilled agricultural trade	.06	.33	6	.01
52 Skilled metal & electrical trade	.23	.67	18	.12
53 Skilled construction & building trades	.03	.24	6	.03
54 Textiles, printing & other skilled trades	.07	.43	12	.09
61 Caring personal service occupations	.03	.45	29	.25
62 Leisure & other personal service occupations	.04	.33	17	.15
71 Sales occupations	.04	.72	18	.36
72 Customer service occupations	.14	.97	-	-
81 Process, plant & machine operatives	.16	.56	15	.10
82 Transport & mobile machine drivers and operatives	.03	.33	9	.05
91 Elementary trades, plant & storage occupations	.03	.42	10	.06
92 Elementary administration & services occupations	.02	.26	8	.12

Source: author's computation based on the 2006th wave of the British Sill Survey (7,787 observations: average advanced computer use at work and use of personal computer (PC)) and the 2006th wave of the German BIBB labour force survey (20,000 observations: percentage time working with a PC). The occupation "Customer service occupations" has no equivalent in the German dataset.

Table A.15 and Table A.16 show the correlation between annual (respectively weekly) wage bill and the proportion of workers having an advanced use of computer, proportion of workers with a university degree and the population of the area (by occupation/TTWA). It shows that the proportion of workers with an advanced use of computer is a relevant dimension for the firms' demand in large cities, and this also controlling for formal education (interaction log advanced PC, log population). Formal education though is also significant (triple interaction).

Table A.15: Local annual wage bill's share and intensity of advanced computer use by occupation in Great Britain

Dep.var. local annual wage bill's share$_{k,j}$	(1)	(2)
VARIABLES	OLS I	OLS II
Log pop$_{k,j}$	.0079***	.0053***
	(.0007)	(.0008)
Log advanced PC$_{k,j}$	-.0372***	-.0330***
	(.0038)	(.0037)
Log advanced PC$_{k,j}$ $\times$ Log pop$_{k,j}$	.0038***	.0029***
	(.0003)	(.0003)
Uni$_{k,j}$		-.0785***
		(.0245))
Uni$_{k,j}$ $\times$ Log pop$_{k,j}$		.0109***
		(.0021)
Uni$_{k,j}$ $\times$ Log advanced PC$_{k,j}$		-.0565***
		(.0175)
Uni$_{k,j}$ $\times$ Log advanced PC$_{k,j}$ $\times$ Log pop$_{k,j}$		.0061***
		(.0015)
Observations	7,128	7,128
R-squared	0.026	0.084

Std. Err. adjusted for 297 clusters (TTWA)

*** p<0.01, ** p<0.05, * p<0.1

Source: author's computation based on the 2006th wave of the Annual Survey of Hours and Earnings as well as the 2006th wave of the British Skill Survey.

Notes: the log of population is the log of population by travel to work area (indexed by . j) which represents an integrated labour market based on commuting flows. Ln advanced computer is the national average of the individual dummy variables of having an advanced computer use at work by SOC2000 occupations at the 2-digits level (indexed by k)

Table A.16: Local weekly wage bill's share and intensity of advanced computer use by occupation in Great Britain

Dep.var. local weekly wage bill's share$_{k,j}$	(1)	(2)
VARIABLES	OLS I	OLS II
Ln pop$_{k,j}$	.0080***	.0058***
	(.0006)	(.0007)
Ln advanced PC$_{k,j}$	-.0390***	-.0360***
	(.0035)	(.0035))
Ln advanced PC$_{k,j}$ $\times$ Ln pop$_{k,j}$	.0038***	.0031***
	(.0003)	(.0003)
Uni$_{k,j}$		-.0688***
		(.0233)
Uni$_{k,j}$ $\times$ Ln pop$_{k,j}$		.0098***
		(.0020)
Uni$_{k,j}$ $\times$ Ln advanced PC$_{k,j}$		-.0539****
		(.0172)
Uni$_{k,j}$ $\times$ Ln advanced PC$_{k,j}$ $\times$ Ln pop$_{k,j}$		.0058***
		(.0015)
Observations	7,128	7,128
R-squared	0.022	0.076

Std. Err. adjusted for 297 clusters (TTWA)

*** p<0.01, ** p<0.05, * p<0.1

Source: Author's computation based on the 2006th wave of the Annual Survey of Hours and Earnings as well as the 2006th wave of the British Skill Survey.

Notes: The log of population is the log of population by travel to work area (indexed by j) which represents an integrated labour market based on commuting flows. Ln advanced computer is the national average of the individual dummy variables of having an advanced computer use at work by SOC2000 occupations at the 2-digits level (indexed by k).

Appendix B

Chapter 2

B.1 Derivation of the proportion of workers in group λ choosing occupation-equipment pair (ω, κ) in city j

- Wage is max in occupation-equipment pairs for a worker if $v_t(\lambda, \kappa, \omega, j)\varepsilon(z, \kappa, \omega) > \max_{\kappa', \omega' \neq \kappa, \omega} \{v_t(\lambda, \kappa', \omega', j)\varepsilon(z, \kappa', \omega')\}$

- The probability that this is the case for a random worker is the probability that her idiosyncratic productivity for occupation ω and equipment κ $\varepsilon(z, \kappa, \omega)$ is larger than $\max_{\kappa', \omega' \neq \kappa, \omega} \left\{ \dfrac{v_t(\lambda, \kappa', \omega', j)\varepsilon(z, \kappa', \omega')}{v_t(\lambda, \kappa, \omega, j)} \right\}$ for all possible values of $\varepsilon(z, \kappa, \omega)$ among workers in group λ is:

$$\pi_t(\lambda, \kappa, \omega \mid j) = \int_0^\infty Pr \left[\varepsilon > \max_{\kappa', \omega' \neq \kappa, \omega} \left\{ \frac{v_t(\lambda, \kappa', \omega', j)\varepsilon(z, \kappa', \omega')}{v_t(\lambda, \kappa, \omega, j)} \right\} \right] dG(\varepsilon \mid j) \quad \text{(B.1)}$$

  where $G(\varepsilon \mid j)$ is the cumulative distribution function of ε given the choice of the city j.

- Because ε are independent across occupation-equipment pairs and the draw of ε is independent of the city j, equation (B.1) is equivalent to:

$$\int_0^\infty \prod_{\kappa', \omega' \neq \kappa, \omega} Pr \left[\varepsilon > \left\{ \frac{v_t(\lambda, \kappa', \omega', j)\varepsilon(z, \kappa', \omega')}{v_t(\lambda, \kappa, \omega, j)} \right\} \right] dG(\varepsilon)$$

$$= \int_0^\infty \exp \left[- \sum_{\kappa', \omega' \neq \kappa, \omega} \left(\frac{\varepsilon v_t(\lambda, \kappa, \omega, j)}{v_t(\lambda, \kappa', \omega', j)} \right)^{-\theta} \right] \theta \varepsilon^{-1-\theta} exp(-\varepsilon^{-\theta}) d\varepsilon$$

  where $G(\varepsilon) = \exp(\varepsilon^{-\theta})$ is the cumulative distribution function of ε and $g(\varepsilon) = \theta\varepsilon^{-1-\theta} exp(-\varepsilon^{-\theta})$ is the probability density function of ε

- Rearranging leads to:

$$= \int_0^\infty \exp -\varepsilon^{-\theta} \left[-\sum_{\kappa',\omega'} \left(\frac{v_t(\lambda,\kappa,\omega,j)}{v_t(\lambda,\kappa',\omega',j)} \right)^{-\theta} \right] \theta \varepsilon^{-1-\theta} d\varepsilon$$

where the sum now includes the κ, ω pair as its first term.

- Substituting the sum by $n_t(\lambda,\kappa,\omega,j)$ and integrating gives:

$$= \int_0^\infty \exp\left(-\varepsilon^{-\theta} n_t(\lambda,\kappa,\omega,j)\right)^{-\theta} (-\theta)\, \varepsilon^{-1-\theta} d\varepsilon$$

$$= \frac{1}{n_t(\lambda,\kappa,\omega,j)} \exp\left(-\varepsilon^{-\theta} n_t(\lambda,\kappa,\omega,j)\right) |_{\varepsilon=0}^\infty$$

$$= \frac{1}{n_t(\lambda,\kappa,\omega,j)}$$

- Substituting back for $n_t(\lambda,\kappa,\omega,j) = \sum_{\kappa',\omega'} \left(\frac{v_t(\lambda,\kappa,\omega,j)}{v_t(\lambda,\kappa',\omega',j)} \right)^{-\theta}$ and $v_t(\lambda,\kappa,\omega,j)$ by equation (2.6), we get equation (2.7).

B.2 Derivation of the average wage of workers in group λ choosing city j

- The average efficiency units ε of workers belonging in group $\lambda \in \mathcal{Z}_t(\lambda) \subseteq \mathcal{Z}_t$ having chosen occupation-equipment (ω,κ) in city j $\bar{\varepsilon}(\lambda,\kappa,\omega,j) = \mathbb{E}\left[\varepsilon(z,\kappa,\omega,j) \mid z \in \mathcal{Z}_t(\lambda,\kappa,\omega,j)\right]$ is equal to [conditional expectation]:

$$\bar{\varepsilon}(\lambda,\kappa,\omega,j) = \frac{1}{\pi_t(\lambda,\kappa,\omega \mid j)} \int_0^\infty \varepsilon \times Pr\left[\varepsilon > \max_{\kappa',\omega' \neq \kappa,\omega} \left\{ \frac{v_t(\lambda,\kappa',\omega',j)\varepsilon(z,\kappa',\omega')}{v_t(\lambda,\kappa,\omega,j)} \right\} \right] dG(\varepsilon \mid j)$$

$$= \frac{1}{\pi_t(\lambda,\kappa,\omega \mid j)} \int_0^\infty \exp -\varepsilon^{-\theta} \left[-\sum_{\kappa',\omega'} \left(\frac{v_t(\lambda,\kappa,\omega,j)}{v_t(\lambda,\kappa',\omega',j)} \right)^{-\theta} \right] \theta \varepsilon^{-\theta} d\varepsilon$$

where the intermediate steps are similar to the proof of Appendix B.1.

- This integral will be solved by substitution (as a function of the Gamma function):

First let $i = \varepsilon^{-\theta}$ [first substitution, $i' = -\theta\varepsilon^{-\theta-1}$] and the sum be represented by $n_t(\lambda,\kappa,\omega,j)$ as before:

$$\bar{\varepsilon}(\lambda,\kappa,\omega,j) = \frac{1}{\pi_t(\lambda,\kappa,\omega \mid j)} \int_\infty^0 \exp\left(-j n_t(\lambda,\kappa,\omega,j)\right) j^{-1/\theta} (-dj)$$

$$= \frac{1}{\pi_t(\lambda, \kappa, \omega \mid j)} \int_0^{\infty} \exp\left(-jn_t(\lambda, \kappa, \omega, j)\right) j^{-1/\theta}(dj)$$

Second, let $y_t(\lambda, \kappa, \omega, j) = n_t(\lambda, \kappa, \omega, j)j$ [second substitution with $y' = n$]:

$$\bar{\varepsilon}(\lambda, \kappa, \omega, j) = \frac{1}{\pi_t(\lambda, \kappa, \omega \mid j)} \int_0^{\infty} \left(\frac{y_t(\lambda, \kappa, \omega, j)}{n_t(\lambda, \kappa, \omega, j)}\right)^{\frac{-1}{\theta}} \exp\left(-y_t(\lambda, \kappa, \omega, j)\right) \frac{dy_t(\lambda, \kappa, \omega, j)}{n_t(\lambda, \kappa, \omega, j)}$$

$$= \frac{1}{\pi_t(\lambda, \kappa, \omega \mid j)} n_t(\lambda, \kappa, \omega, j)^{\frac{1-\theta}{\theta}} \int_0^{\infty} \left(y_t(\lambda, \kappa, \omega, j)\right)^{\frac{-1}{\theta}} \exp\left(-y_t(\lambda, \kappa, \omega, j)\right) dy_t(\lambda, \kappa, \omega, j)$$

$$= \frac{1}{\pi_t(\lambda, \kappa, \omega \mid j)} n_t(\lambda, \kappa, \omega, j)^{\frac{1-\theta}{\theta}} \times \gamma$$

where $\gamma \equiv \Gamma(1 - \frac{1}{\theta})$ and $\Gamma(.)$ is the Gamma function:

$$\Gamma(x) \equiv \int_0^{\infty} t^{x-1} \exp(-t) dt$$

- Inserting $n_t(\lambda, \kappa, \omega, j) =$ back into the equation leads to:

$$\bar{\varepsilon}(\lambda, \kappa, \omega, j) = \gamma \times \pi_t(\lambda, \kappa, \omega \mid j)^{\frac{-1}{\theta}}$$

- Now the average wage rate $w_t(\lambda, \kappa, \omega, j)$ is $\bar{\varepsilon}(\lambda, \kappa, \omega, j)v_t(\lambda, \kappa, \omega, j)$
- Substituting for $v_t(\lambda, \kappa, \omega, j)$ with equation (2.6), then $\pi_t(\lambda, \kappa, \omega, j)$ with equation (2.7) leads to:

$$w_t(\lambda, j) = \gamma(\theta)\bar{\alpha} p_t(\kappa)^{-\frac{\alpha}{1-\alpha}} p_t(\omega)^{\frac{1}{1-\alpha}} T_t(\lambda, \kappa, \omega, j) \pi_t(\lambda, \kappa, \omega, j)^{\frac{-1}{\theta}}$$

$$= \gamma(\theta)\bar{\alpha} p_t(\kappa)^{-\frac{\alpha}{1-\alpha}} p_t(\omega)^{\frac{1}{1-\alpha}} T_t(\lambda, \kappa, \omega, j) \frac{\left[T_t(\lambda, \kappa, \omega, j) p_t(\kappa)^{-\frac{\alpha}{1-\alpha}} p_t(\omega)^{\frac{1}{1-\alpha}}\right]^{\theta}}{\sum_{\kappa', \omega'} \left[T_t(\lambda, \kappa', \omega', j) p_t(\kappa')^{-\frac{\alpha}{1-\alpha}} p_t(\omega')^{\frac{1}{1-\alpha}}\right]^{\theta}}^{\frac{-1}{\theta}}$$

$$w_t(\lambda, j) = \gamma(\theta)\bar{\alpha} \left\{ \sum_{\kappa', \omega'} \left[T_t(\lambda, \kappa', \omega', j) p_t(\kappa')^{-\frac{\alpha}{1-\alpha}} p_t(\omega')^{\frac{1}{1-\alpha}}\right]^{\theta} \right\}^{\frac{1}{\theta}}$$

which is equation (2.8)

B.3 Derivation of the proportion of workers in group λ choosing city j

- Expected indirect utility of a worker in city j is max if: $\frac{w_t(\lambda,j)Q_t(\lambda,j)^{\varrho(\lambda)}\psi(z,\lambda)}{p_t(j)^{\beta(\lambda)}} > \max_{j'\neq j}\left\{\frac{w_t(\lambda,j')Q_t(\lambda,j')\psi(z,j')}{p_t(j')^{\beta(\lambda)}}\right\}$
- The probability that this is the case for a worker z in group $\lambda \in \mathcal{Z}_t(\lambda) \subseteq \mathcal{Z}_t$ is

$$\xi_t(\lambda,j) = \int_0^\infty Pr\left[\psi > \max_{\kappa',\omega'\neq\kappa,\omega}\left\{\frac{w_t(\lambda,j')Q_t(\lambda,j')\psi(z,j')}{w_t(\lambda,j)Q_t(\lambda,j)^{\varrho(\lambda)}}\frac{p_t(j)^{\beta(\lambda)}}{p_t(j')^{\beta(\lambda)}}\right\}\right] dG(\psi) \quad \text{(B.2)}$$

  where $G(\psi) =$ is the cumulative distribution function of ψ.
- Further steps are analogous to deriving equation (2.7)

B.4 Proof of Lemma 4

To prove Lemma 4, I introduce the following notation, which will turn easier to prove differentiability, existence and uniqueness.

Notation in terms of the $\tilde{p}_t(j) = \log p_t(j)$, $\tilde{p}_t(\omega) = \log p_t(\omega)$ and $\tilde{w}_t(\lambda,j) = \log w_t(\lambda,j)$:

$$\begin{aligned}\tilde{p}_t(j) =& \frac{1}{1+2\beta\eta}\left[\log\left(\sum_\lambda Q_t(\lambda,j)^{\varrho(\lambda)\eta} e^{(1+\eta)\tilde{w}_t(\lambda,j)-\log\left(\sum_{j'} Q_t(\lambda,j')^\eta \exp\{\eta\tilde{w}_t(\lambda,j')-\beta\eta\tilde{p}_t(j')\}\right)}\right)\right. \\ &\left. + \log\left(\sum_\lambda L_t(\lambda)Q_t(\lambda,j)^{\varrho(\lambda)\eta} e^{\eta\tilde{w}_t(\lambda,j)-\log\left(\sum_{j'} Q_t(\lambda,j')^\eta \exp\{\eta\tilde{w}_t(\lambda,j')-\beta\eta\tilde{p}_t(j')\}\right)}\right)\right]\end{aligned} \quad \text{(B.3)}$$

where $\tilde{w}_t(\lambda,j) = \log(\gamma(\theta)\bar{\alpha}) + \frac{1}{\theta}\log\left(\sum_{\kappa',\omega'}\exp\left\{\theta\log\left(T_t(\lambda,\kappa',\omega',j)p_t(\kappa')^{-\frac{\alpha}{1-\alpha}}\right) + \frac{1}{1-\alpha}\theta\tilde{p}_t(\omega')\right\}\right)$

$j = 1,...,m$ and $\omega = 1,...,n$. Define the function $f : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^m$ so that these n equations are:

$$\tilde{p}_t(j) = f_t(j, \boldsymbol{\tilde{p}_t(j)}, \boldsymbol{\tilde{p}_t(\omega)}) \quad \text{(B.4)}$$

The derivative of equation (B.4) with respect to the price in one given city, denoted j' is given by:

$\frac{\partial f_t(j,\boldsymbol{\tilde{p}_t(j)},\boldsymbol{\tilde{p}_t(\omega)})}{\partial\tilde{p}_t(j')} = \frac{\beta\eta}{1+2\beta\eta}\chi_{jj'}$

where $\chi_{jj'}$ is given by :

$$\chi_{jj'} = \frac{\sum_\lambda \left(\frac{Q_t(\lambda,j)^{\varrho(\lambda)\eta} w(\lambda,j)^{1+\eta} Q_t(\lambda,j')^\eta w(\lambda,j')^\eta p_t(j')^{-\beta\eta}}{\left[\sum_{j''=1}^m Q_t(\lambda,j'')^\eta w(\lambda,j'')^\eta p_t(j'')^{-\beta\eta}\right]^2} \right)}{\sum_\lambda \left(\frac{Q_t(\lambda,j)^{\varrho(\lambda)\eta} w(\lambda,j)^{1+\eta}}{\sum_{j''}\left[\frac{w_t(\lambda,j'')Q_t(\lambda,j'')}{p_t(j'')^\beta}\right]^\eta} \right)} + \frac{\sum_\lambda \left(\frac{L_t(\lambda) Q_t(\lambda,j)^{\varrho(\lambda)\eta} w(\lambda,j)^{\eta} Q_t(\lambda,j')^\eta w(\lambda,j')^\eta p_t(j')^{-\beta\eta}}{\left[\sum_{j''=1}^m Q_t(\lambda,j'')^\eta w(\lambda,j'')^\eta p_t(j'')^{-\beta\eta}\right]^2} \right)}{\sum_\lambda \left(\frac{L_t(\lambda) Q_t(\lambda,j)^{\varrho(\lambda)\eta} w(\lambda,j)^{\eta}}{\sum_{j''}\left[\frac{w_t(\lambda,j'')Q_t(\lambda,j'')}{p_t(j'')^\beta}\right]^\eta} \right)} \tag{B.5}$$

Let $S_1 = [\chi_{jj'}]$ be the $m \times m$ matrix with elements $\chi_{jj'}$.

Next, I turn to the derivative of the log of the price in cities with respect to the price of one occupation, denoted ω'. Note that it appears in the wage of every skill group λ and in every city j. As before, bold letters denote vectors, e.g. $\boldsymbol{\tilde{w}_t(\lambda, j)} = (\tilde{w}_t(\lambda, 1), ..., \tilde{w}_t(\lambda, m))$.

$\frac{\partial f_t(j, \boldsymbol{\tilde{p}_t(j)}, \boldsymbol{\tilde{p}_t(\omega)})}{\partial \tilde{p}_t(\omega')} = \frac{\partial f_t(j, \boldsymbol{\tilde{p}_t(j)}, \boldsymbol{\tilde{p}_t(\omega)})}{\partial \boldsymbol{\tilde{w}_t(\lambda, j)}} \frac{\partial \boldsymbol{\tilde{w}_t(\lambda, j)}}{\partial \tilde{p}_t(\omega')} = \frac{1}{1-\alpha}\frac{1}{1+2\beta\eta}\varsigma_{j\omega'}$

where $\varsigma_{j\omega'}$ is given by:

$$\varsigma_{j\omega'} = \frac{\sum_{\kappa'} \left[T_t(\lambda, \kappa', \omega', j) p_t(\kappa')^{-\frac{\alpha}{1-\alpha}} p_t(\omega')^{\frac{1}{1-\alpha}}\right]^\theta}{\sum_{\kappa', \omega''} \left[T_t(\lambda, \kappa', \omega'', j) p_t(\kappa')^{-\frac{\alpha}{1-\alpha}} p_t(\omega'')^{\frac{1}{1-\alpha}}\right]^\theta} \tag{B.6}$$

$$\frac{\partial \tilde{w}_t(\lambda, j)}{\partial \tilde{p}_t(\omega')} = \frac{1}{1-\alpha} \frac{\sum_{\kappa'} \left[T_t(\lambda, \kappa', \omega', j) p_t(\kappa')^{-\frac{\alpha}{1-\alpha}} p_t(\omega')^{\frac{1}{1-\alpha}}\right]^\theta}{\sum_{\kappa', \omega''} \left[T_t(\lambda, \kappa', \omega'', j) p_t(\kappa')^{-\frac{\alpha}{1-\alpha}} p_t(\omega'')^{\frac{1}{1-\alpha}}\right]^\theta} \tag{B.7}$$

$$\frac{\partial \tilde{p}_t(j)}{\partial \boldsymbol{\tilde{w}_t(\lambda, j)}} = \frac{1}{1+2\beta\eta} \frac{\sum_\lambda \left(\frac{Q_t(\lambda,j)^{\varrho(\lambda)\eta} w(\lambda,j)^{1+\eta}}{\sum_{j'}\left[\frac{w_t(\lambda,j')Q_t(\lambda,j')}{p_t(j')^\beta}\right]^\eta} \left[(1+\eta) - \eta \frac{\sum_{j'=1}^m Q_t(\lambda,j')^\eta w_t(\lambda,j')^\eta p_t(j')^{-\beta\eta}}{\sum_{j'=1}^m Q_t(\lambda,j')^\eta w_t(\lambda,j')^\eta p_t(j')^{-\beta\eta}}\right] \right)}{\sum_\lambda \left(\frac{Q_t(\lambda,j)^{\varrho(\lambda)\eta} w_t(\lambda,j)^{1+\eta}}{\sum_{j'}\left[\frac{w_t(\lambda,j')Q_t(\lambda,j')}{p_t(j')^\beta}\right]^\eta} \right)}$$

$$+ \frac{\eta}{1+2\beta\eta} \frac{\sum_\lambda \left(\frac{L_t(\lambda) Q_t(\lambda,j)^{\varrho(\lambda)\eta} w_t(\lambda,j)^{1+\eta}}{\sum_{j'}\left[\frac{w_t(\lambda,j')Q_t(\lambda,j')}{p_t(j')^\beta}\right]^\eta} \left[1 - \frac{\sum_{j'=1}^m Q_t(\lambda,j')^\eta w_t(\lambda,j')^\eta p_t(j')^{-\beta\eta}}{\sum_{j'=1}^m Q_t(\lambda,j')^\eta w_t(\lambda,j')^\eta p_t(j')^{-\beta\eta}}\right] \right)}{\sum_\lambda \left(\frac{L_t(\lambda) Q_t(\lambda,j)^{\varrho(\lambda)\eta} w_t(\lambda,j)^{1+\eta}}{\sum_{j'}\left[\frac{w_t(\lambda,j')Q_t(\lambda,j')}{p_t(j')^\beta}\right]^\eta} \right)}$$

$$\frac{\partial \tilde{p}_t(j)}{\partial \boldsymbol{\tilde{w}_t(\lambda, j)}} = \frac{1}{1+2\beta\eta}$$

Let $S_2 = [\varsigma_{j\omega'}]$ be the $m \times n$ matrix with elements $\varsigma_{j\omega'}$. Note that S_2 is a strictly positive matrix.

The jacobian of the system $\tilde{p}_t(j) - f_t(j, \boldsymbol{\tilde{p}_t(j)}, \boldsymbol{\tilde{p}_t(\omega)})$ with respect to $\tilde{p}_t(j)$ is $I - \frac{\beta\eta}{1+2\beta\eta} S_1$. $\chi_{jj'} > 0$ and $\chi_{jj'} < 2$. $\frac{\beta\eta}{1+2\beta\eta} < \frac{1}{2}$, so that the inverse of the Jacobian is a strictly positive matrix (the series below converges):

$$\left[I - \frac{\beta\eta}{1+2\beta\eta} S_1 \right]^{-1} = \sum_{i=0}^{\infty} \left(\frac{\beta\eta}{1+2\beta\eta} \right)^i S_1^i \tag{B.8}$$

Part I and III

The implicit function theorem implies that $\frac{\partial \tilde{p}_t(j)}{\partial \tilde{p}_t(\omega')}$ is well defined for any $p_t(\omega) \in \mathbf{R}^n_{++}$ and given by:

$$\frac{\partial \tilde{p}_t(j)}{\partial \tilde{p}_t(\omega')} = \left[I - \frac{\beta\eta}{1+2\beta\eta} S_1 \right]^{-1} \frac{1}{1-\alpha} \frac{1}{1+2\beta\eta} \varsigma_{j\omega'} \tag{B.9}$$

This derivative is positive $\forall\ p_t(\omega)$ and $p_t(j) \in \mathbf{R}^n_{++}$.

Part II

The wage is homogeneous of degree 1 in prices because of constant returns to scale. Then by multiplying the wage and city price by a, the city price on the left hand side indeed is multiplied by a:

$$a p_t(j) = \left\{ \sum_{\lambda} \frac{a^{1+\eta} w_t(\lambda, j)^{1+\eta} Q_t(\lambda, j)^{\varrho(\lambda)\eta}}{\sum_{j'} \left[\frac{a w_t(\lambda, j') Q_t(\lambda, j')}{a^{\beta} p_t(j')^{\beta}} \right]^{\eta}} \sum_{\lambda} \frac{L_t(\lambda) a^{\eta} w_t(\lambda, j)^{\eta} Q_t(\lambda, j)^{\varrho(\lambda)\eta}}{\sum_{j'} \left[\frac{a w_t(\lambda, j') Q_t(\lambda, j')}{a^{\beta} p_t(j')^{\beta}} \right]^{\eta}} \right\}^{\frac{1}{1+2\beta\eta}}$$

Note that homothetic preferences are crucial for this result. Indeed, there is no wealth effect, so that if the wage is multiplied by x for all household in the city, the city price also will be multiplied by x. [1]

[1](pour moi, revois la fonction d'utilité indirecte pour des préférences non-homothetic, pour comparer, mets-le en commmentaires dans ce mansucript)

Part IV

The parameters $T_t(\lambda, \kappa, \omega, j)$ and $Q_t(\lambda, j)^{\varrho(\lambda)}$ all increase the attractivity of city j relative to other cities, and thus its price $p_t(j)$. First, it increases the wage in city j and city price through that channel (equation (2.10)). Second, it increases city population by raising the probability that workers choose to live and work in that city (equation (2.9)) and thus city price. The parameters $T_t(\lambda, \kappa, \omega, j')$ and $Q_t(\lambda, j')$, which refer to city $j' \neq j$, decrease the attractivity of city j relative to city j' and thus decrease its price $p_t(j)$.

Part V

In order to derive the bounds, I consider the case where one city denoted j has the characteristics $w_t(j)$ and $Q_t(j)$ and all other cities denotes j'' has the characteristics $w_t(j'')$ and $Q_t(j'')$. In this case, equation (2.11) becomes, where Λ denotes the number of skill group λ:

$$p_t(j) = \left\{ \frac{\Lambda^2 L_t w_t(j)^{1+2\eta} Q_t(j)^{2\eta}}{\left[\frac{w_t(j)^\eta Q_t(j)^\eta}{p_t(j)^{\beta\eta}} + (m-1)\frac{w_t(j'')^\eta Q_t(j'')^\eta}{p_t(j'')^{\beta\eta}}\right]^2} \right\}^{\frac{1}{1+2\beta\eta}}$$

Isolating $p_t(j'')$ from this equation gives:

$$p_t(j'') = \left\{ \frac{(m-1) w_t(j'')^\eta Q_t(j'')^\eta p_t(j)^{\frac{1+2\beta\eta}{2}}}{\Lambda L_t^{\frac{1}{2}} w_t(j)^{\frac{1+2\eta}{2}} Q_t(j)^\eta - w_t(j)^\eta Q_t(j)^\eta p_t(j)^{\frac{1}{2}}} \right\}^{\frac{1}{\beta\eta}} \tag{B.10}$$

Second, the price in cities j" is equal to:

$$p_t(j'') = \left\{ \frac{\Lambda^2 L_t w_t(j'')^{1+2\eta} Q_t(j'')^{2\eta}}{\left[\frac{w_t(j)^\eta Q_t(j)^\eta}{p_t(j)^{\beta\eta}} + (m-1)\frac{w_t(j'')^\eta Q_t(j'')^\eta}{p_t(j'')^{\beta\eta}}\right]^2} \right\}^{\frac{1}{1+2\beta\eta}} \tag{B.11}$$

Plugging equation (B.10) in equation (B.11) gives:

$$\left\{\frac{(m-1)w_t(j'')^\eta Q_t(j'')^\eta p_t(j)^{\frac{1+2\beta\eta}{2}}}{\Lambda L_t^{\frac{1}{2}} w_t(j)^{\frac{1+2\eta}{2}} Q_t(j)^\eta - w_t(j)^\eta Q_t(j)^\eta p_t(j)^{\frac{1}{2}}}\right\}^{\frac{1}{\beta\eta}}$$
$$= \left\{\frac{\Lambda^2 L_t w_t(j'')^{1+2\eta} Q_t(j'')^{2\eta}}{\left[\frac{w_t(j)^\eta Q_t(j)^\eta}{p_t(j)^{\beta\eta}} + \frac{(m-1)w_t(j'')^\eta Q_t(j'')^\eta}{\frac{(m-1)w_t(j'')^\eta Q_t(j'')^\eta p_t(j)^{\frac{1+2\beta\eta}{2}}}{\Lambda L_t^{\frac{1}{2}} w_t(j)^{\frac{1+2\eta}{2}} Q_t(j)^\eta - w_t(j)^\eta Q_t(j)^\eta p_t(j)^{\frac{1}{2}}}}\right]^2}\right\}^{\frac{1}{1+2\beta\eta}} \tag{B.12}$$

Simplifying to isolate $p_t(j)$ gives:

$$p_t(j) = \Lambda^2 L_t \left[\frac{(m-1)w_t(j'')^{\frac{\eta(1-\beta)}{1+2\beta\eta}} Q_t(j'')^{\frac{\eta(1+\beta\eta)}{1+2\beta\eta}}}{w_t(j)^{\frac{1+2\eta}{2(1+2\beta\eta)}} Q_t(j)^{\frac{\eta(1+\beta\eta)}{1+2\beta\eta}}} + w_t(j)^{-1/2}\right]^{-2} \tag{B.13}$$

To get to that expression, I first simplify the denominator of the RHS of equation (B.12) which gives equation (B.14) below. I then take the expression to the power of $1/(\beta\eta)$ to be able to isolate $p_t(j)$.

$$\left\{\frac{(m-1)w_t(j'')^\eta Q_t(j'')^\eta p_t(j)^{\frac{1+2\beta\eta}{2}}}{\Lambda L_t^{\frac{1}{2}} w_t(j)^{\frac{1+2\eta}{2}} Q_t(j)^\eta - w_t(j)^\eta Q_t(j)^\eta p_t(j)^{\frac{1}{2}}}\right\}^{\frac{1}{\beta\eta}}$$
$$= \left\{\frac{p_t(j)^{1+2\beta\eta}\Lambda^2 L_t w_t(j'')^{1+2\eta} Q_t(j'')^{2\eta}}{\left[\Lambda L_t^{\frac{1}{2}} w_t(j)^{\frac{1+2\eta}{2}} Q_t(j)^\eta\right]^2}\right\}^{\frac{1}{1+2\beta\eta}} \tag{B.14}$$

The upper bound $\overline{p}(j, \boldsymbol{p_t(\omega)})$ results from one city having the highest productivity $T_t(.) = 1$ and amenity $Q_t(.) = 1$, while all other cities have the lowest productivity $\underline{T}$ and amenity $\underline{Q}$. This is the case because of the monotonicity properties of Lemma 4 part (III) and (IV). Equation (B.13) thus become:

$$\overline{p}(j, \boldsymbol{p_t(\omega)}) = \Lambda^2 L_t \left[\frac{(m-1)\underline{w}(\boldsymbol{p_t(\omega)})^{\frac{\eta(1-\beta)}{1+2\beta\eta}} \underline{Q}^{\frac{\eta(1+\beta\eta)}{1+2\beta\eta}}}{\overline{w}(\boldsymbol{p_t(\omega)})^{\frac{1+2\eta}{2(1+2\beta\eta)}}} + \overline{w}(\boldsymbol{p_t(\omega)})^{-1/2}\right]^{-2} \tag{B.15}$$

where $\underline{w}(\boldsymbol{p_t(\omega)}) = \gamma(\theta)\bar{\alpha}\left\{\sum_{\kappa',\omega'}\left[\underline{T}p_t(\kappa')^{-\frac{\alpha}{1-\alpha}}p_t(\omega')^{\frac{1}{1-\alpha}}\right]^{\theta}\right\}^{\frac{1}{\theta}}$
and $\overline{w}(\boldsymbol{p_t(\omega)}) = \gamma(\theta)\bar{\alpha}\left\{\sum_{\kappa',\omega'}\left[p_t(\kappa')^{-\frac{\alpha}{1-\alpha}}p_t(\omega')^{\frac{1}{1-\alpha}}\right]^{\theta}\right\}^{\frac{1}{\theta}}$

The lower bound $\underline{p}(j, \boldsymbol{p_t(\omega)})$ results from one city having the lowest productivity $\underline{T}$ and amenity $\underline{Q}$, while all other cities have the highest values $T_t(.) = 1$ and $Q_t(.) = 1$. Equation (B.13) thus become:

$$\underline{p}(j, \boldsymbol{p_t(\omega)}) = \Lambda^2 L_t\left[\frac{(m-1)\overline{w}(\boldsymbol{p_t(\omega)})^{\frac{\eta(1-\beta)}{1+2\beta\eta}}}{\underline{w}(\boldsymbol{p_t(\omega)})^{\frac{1+2\eta}{2(1+2\beta\eta)}}\underline{Q}^{\frac{\eta(1+\beta\eta)}{1+2\beta\eta}}} + \underline{w}(\boldsymbol{p_t(\omega)})^{-1/2}\right]^{-2} \tag{B.16}$$

I next complete the proof of Lemma 4:

The bounds (V), with the monotonicity properties (III) and (IV), ensure that any solution to equation (B.3) must satisfy the bounds.

Definition of the set of solutions: for a given $\boldsymbol{p_t(\omega)} \in \mathbf{R}^n_{++}$ the set $\mathbf{C}$:

$\mathbf{C} = \left\{z \in \mathbf{R}^n : \log \underline{p}(j, \boldsymbol{p_t(\omega)}) \leqslant z_t(j) \leqslant \log \overline{p}(j, \boldsymbol{p_t(\omega)}),\ \forall j\right\}$

Under the sup norm $\|z\| = \max_j |z_t(j)|$, $\mathbf{C}$ is compact.

The function $f_t(j, \boldsymbol{\tilde{p}_t(j)}, \boldsymbol{\tilde{p}_t(\omega)})$ maps itself into the set $\mathbf{C}$: $f_t(j, \boldsymbol{\tilde{p}_t(j)}, \boldsymbol{\tilde{p}_t(\omega)}) : \mathbf{C} \to \mathbf{C}$:

Using the bounds on $Q_t(\lambda, j)$ and $T_t(\lambda, \kappa', \omega', j)$ (and thus on $p_t(j)$), the monotonicity properties and $f_t(z, \boldsymbol{\tilde{p}_t(\omega)}$ increasing in z (if the vector of city prices increases on the right hand side, consistent with an increase of the vector of city prices on the left hand side since all city prices positively depends on city prices in other cities), then for any $z \in \mathbf{C}$, Q, T and $p_t(\omega)$:

$f_t(z, \boldsymbol{\tilde{p}_t(\omega)}, \boldsymbol{Q}, \boldsymbol{T}) \leqslant \log \overline{p}(j, \boldsymbol{p_t(\omega)}) = f(\log \overline{p}(j, \boldsymbol{p_t(\omega)}), \boldsymbol{\tilde{p}_t(\omega)}, Q(j) = 1, T(j) = 1, Q(j') = \underline{Q}, T(j') = \underline{T})$, $j \neq j'$

And $f_t(z, \boldsymbol{\tilde{p}_t(\omega)}, \boldsymbol{Q}, \boldsymbol{T}) \geqslant \log \underline{p}(j, \boldsymbol{p_t(\omega)}) = f(\log \underline{p}(j, \boldsymbol{p_t(\omega)}), \boldsymbol{\tilde{p}_t(\omega)}, Q(j) = \underline{Q}, T(j) = \underline{T}, Q(j') = 1, T(j') = 1)$, $j \neq j'$.

Last I check the Blackwell sufficient conditions for $f_t(z, \boldsymbol{\tilde{p}_t(\omega)}, \boldsymbol{Q}, \boldsymbol{T})$ to be a contraction.

(i) Monotone property. Already shown

(ii) Discounting property : $f_t(j, z + a) = f_t(j, z) + \gamma a$, $a > 0$, $\gamma \in (0, 1)$

$$
\begin{aligned}
f_t(j, z+a) =& \frac{1}{1+2\beta\eta}\Bigg[\log\left(\sum_\lambda Q_t(\lambda, j)^{\varrho(\lambda)\eta} e^{(1+\eta)\tilde{w}(\lambda,j)-\log\left(\sum_{j'} Q_t(\lambda,j')^\eta \exp\{\eta\tilde{w}(\lambda,j')-\beta\eta(\tilde{p}j'+a)\}\right)}\right) \\
&+ \log\left(\sum_\lambda L_t(\lambda) Q_t(\lambda, j)^{\varrho(\lambda)\eta} e^{\eta\tilde{w}(\lambda,j)-\log\left(\sum_{j'} Q_t(\lambda,j')^\eta \exp\{\eta\tilde{w}(\lambda,j')-\beta\eta(\tilde{p}j+a)\}\right)}\right)\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
=& \frac{1}{1+2\beta\eta}\Bigg[\log\left(\sum_\lambda Q_t(\lambda, j)^{\varrho(\lambda)\eta} e^{(1+\eta)\tilde{w}(\lambda,j)-\log\left(\sum_{j'} Q_t(\lambda,j')^\eta \exp\{\eta\tilde{w}(\lambda,j')-\beta\eta\tilde{p}j'\}\right)}\right) + \beta\eta a \\
&+ \log\left(\sum_\lambda L_t(\lambda) Q_t(\lambda, j)^{\varrho(\lambda)\eta} e^{\eta\tilde{w}(\lambda,j)-\log\left(\sum_{j'} Q_t(\lambda,j')^\eta \exp\{\eta\tilde{w}(\lambda,j')-\beta\eta\tilde{p}j'\}\right)}\right) + \beta\eta a\Bigg]
\end{aligned}
$$

$$
f_t(j, z+a) = f_t(j, z) + a\frac{2\beta\eta}{1+2\beta\eta}
$$

with $\gamma = \frac{2\beta\eta}{1+2\beta\eta} < 1$, using the properties of the exponential and logarithmic functions.

The contraction mapping theorem thus implies the existence of a unique fixed point $\tilde{p}_t(j)$ for $f_t(j, \boldsymbol{\tilde{p}_t(j)}, \boldsymbol{\tilde{p}_t(\omega)})$ and a unique solution $p_t(j)$ for equation (2.11).

B.5 Proof of Proposition 3

Part I

The continuity of $p_t(j)$ is implied by Lemma 4 [I]. The continuity of $Z(\omega, p(\omega))$ then follows from equation (2.14), 2.9, 2.8 and 2.7.

Part II

From Lemma 4 [II], $p_t(j)$ is homogeneous in prices. Then, from equation (2.9), $\xi_t(\lambda, j)$ is homogeneous of degree 0, from equation (2.8), $w_t(\lambda, j)$ is homogeneous of degree 1, thus E_t and $\zeta_t(\omega)$ are homogeneous of degree 1. Last, $\frac{P_t(p_t(\omega))}{p_t(\omega)}$ is homogeneous of degree 0, so that $Z(\omega, p(\omega))$ is homogeneous of degree 0.

Part III

Walras's law: In my case: non profit condition from the final firm. Proof:

$$
p_t(\omega) Z_t(\omega, p(\omega)) = E_t p_t(\omega)^{1-\rho} P_t^{\rho-1} \phi_t(\omega) - \frac{1}{1-\alpha}\zeta_t(\omega) \tag{B.17}
$$

Summing over all occupation ω:

$$= E_t P_t^{\rho-1} \sum_{\omega} p_t(\omega)^{1-\rho} \phi_t(\omega) - E_t = 0 \tag{B.18}$$

since $P_t^{\rho-1} \sum_{\omega} p_t(\omega)^{1-\rho} \phi_t(\omega) = 1$ by the definition of the marginal cost of the final firms.

Part IV

To derive a bound on $Z_t(\omega, p_t(\omega))$, I take the case where production in the economy $Y_t(\omega, j) = k^{\alpha} \left[T_t(\lambda, \kappa, \omega, j)l\right]^{1-\alpha}$, with $k_t(.) = \left(\alpha \frac{p_t(\omega)}{p_t(\kappa)}\right)^{\frac{1}{1-\alpha}} T_t(\lambda, \kappa, \omega, j)l$ result from a single occupation, with the highest productivity: $Y_t(\max \omega) = \left(\alpha \frac{p_t(\omega)}{p_t(\kappa)}\right)^{\frac{\alpha}{1-\alpha}} \max \lambda, \kappa, j T_t(\lambda, \kappa, \omega, j) L_t$. As $p_t(\omega)$ is endogenous in this expression, I derive a bound on it which result from the case described here.

Bound on $p_t(\omega)$:

All workers are employed in the occupation with the highest productivity so that $\xi_t(\lambda, j) = 1$, $\pi_t(\lambda, \omega, \kappa, j) = 1$, $E_t = \dfrac{1}{1-\alpha}\zeta_t(\omega)$ as the labor income in the occupation is also the labor income in the economy. In this case, equation (2.14) set to zero has a closed form solution:

$$Z_t(\omega, p_t(\omega)) = \tfrac{1}{p_t(\omega)} E_t \left[p_t(\omega)^{1-\rho} P_t^{\rho-1} \phi_t(\omega) - 1\right] = 0$$

$$\Leftrightarrow p_t(\omega)^{\rho-1} = \phi_t(\omega)$$
$$\Leftrightarrow p_t(\omega) = \phi_t(\omega)^{\frac{1}{\rho-1}} = 1, \text{ with } \max \phi_t(\omega) = 1.$$

A bound on $Z_t(\omega, p_t(\omega))$ thus is $-\left(\frac{\alpha}{p_t(\kappa)}\right)^{\frac{\alpha}{1-\alpha}} L_t$ as the demand part of the excess demand function will always be bounded ($p_t(\omega) \gg 0$), no free lunch as the final firm positively values the output of the occupation (monotonicity of "preferences").

Part V

If one of the occupation price is zero and all other prices are strictly positive, then it's excess demand $Z_t(\omega, p_t(\omega))$ will tend to infinity. Indeed, its demand will tend to infinity and the supply $\dfrac{1}{(1-\alpha)p_t(\omega)}\zeta_t(\omega)$ will tend to zero as $\frac{\theta-1+\alpha}{1-\alpha} > 0$.

$$Z_t(\omega, p_t(\omega)) = \tfrac{1}{p_t(\omega)} \left[E_t p_t(\omega)^{1-\rho} P_t^{\rho-1} \phi_t(\omega) - \frac{1}{1-\alpha}\zeta_t(\omega)\right]$$

where $E_t = \dfrac{1}{1-\alpha} \sum_{j,\lambda} w_t(j, \lambda) L_t(\lambda) \xi_t(\lambda, j)$, which is in the limit strictly positive as is $w_t(\lambda, j) = \gamma(\theta)\bar{\alpha} \left\{\sum_{\kappa', \omega'} \left[T_t(\lambda, \kappa', \omega', j) p_t(\kappa')^{-\frac{\alpha}{1-\alpha}} p_t(\omega')^{\frac{1}{1-\alpha}}\right]^{\theta}\right\}^{\frac{1}{\theta}}$ and $\xi_t(\lambda, j) = \frac{\left[\frac{w_t(\lambda, j) Q_t(\lambda, j)^{\varrho(\lambda)}}{p_t(j)^{\beta(\lambda)}}\right]^{\eta}}{\sum_{j'} \left[\frac{w_t(\lambda, j') Q_t(\lambda, j')}{p_t(j')^{\beta(\lambda)}}\right]^{\eta}}$.

$\zeta_t(\omega) = \sum_{j,\lambda,\kappa} w_t(j,\lambda)L_t(\lambda)\xi_t(\lambda,j)\pi_t(\lambda,\kappa,\omega,j)$ is the labor income in occupation ω, which limits tends to zero as the probability that workers are employed in that occupation $\pi_t(\lambda,\kappa,\omega \mid j)$ is zero if its price tends to zero:

$$\pi_t(\lambda,\kappa,\omega \mid j) = \frac{\left[T_t(\lambda,\kappa,\omega,j)p_t(\kappa)^{-\frac{\alpha}{1-\alpha}}p_t(\omega)^{\frac{1}{1-\alpha}}\right]^{\theta}}{\sum_{\kappa',\omega'}\left[T_t(\lambda,\kappa',\omega',j)p_t(\kappa')^{-\frac{\alpha}{1-\alpha}}p_t(\omega')^{\frac{1}{1-\alpha}}\right]^{\theta}}.$$

B.6 Proof of Proposition 4

The gross substitute property implies that $\frac{\partial Z_t(\omega,\boldsymbol{p_t(\omega)})}{\partial p_t(\omega'')} > 0 \; \forall\omega,\ \omega \neq \omega'',\ \forall\, p_t(\omega) \in \mathbf{R}^n_{++}$:

$$\begin{aligned}\frac{\partial Z_t(\omega,\boldsymbol{p_t(\omega)})}{\partial p_t(\omega'')} &= \frac{1}{(1-\alpha)p_t(\omega)}\Bigg\{\sum_{j,\lambda} p_t(\omega)^{1-\rho}\phi_t(\omega)L_t(\lambda)\left[\frac{\partial w_t(j,\lambda)}{\partial p_t(\omega'')}\xi_t(\lambda,j) + w_t(j,\lambda)\frac{\partial \xi_t(\lambda,j)}{\partial p_t(\omega'')}\right] \\ &- \sum_{j,\lambda,\kappa} L_t(\lambda)\Bigg[\frac{\partial w_t(j,\lambda)}{\partial p_t(\omega'')}\xi_t(\lambda,j)\pi_t(\lambda,\kappa,\omega,j) + w_t(j,\lambda)\frac{\partial \xi_t(\lambda,j)}{\partial p_t(\omega'')}\pi_t(\lambda,\kappa,\omega,j) \\ &+ w_t(j,\lambda)\xi_t(\lambda,j)\frac{\partial \pi_t(\lambda,\kappa,\omega,j)}{\partial p_t(\omega'')}\Bigg]\Bigg\}\end{aligned} \tag{B.19}$$

$$\frac{\partial Z_t(\omega,\boldsymbol{p_t(\omega)})}{\partial p_t(\omega'')} = \frac{-w_t(j,\lambda)L_t(\lambda)\xi_t(\lambda,j)}{(1-\alpha)p_t(\omega)}\frac{\partial \pi_t(\lambda,\kappa,\omega,j)}{\partial p_t(\omega'')} > 0 \tag{B.20}$$

since $\sum_{j,\lambda} p_t(\omega)^{1-\rho}\phi_t(\omega)L_t(\lambda)\frac{\partial w_t(j,\lambda)}{\partial p_t(\omega'')}\xi_t(\lambda,j) = \sum_{j,\lambda,\kappa} L_t(\lambda)\frac{\partial w_t(j,\lambda)}{\partial p_t(\omega'')}\xi_t(\lambda,j)\pi_t(\lambda,\kappa,\omega,j)$ and $\sum_{j,\lambda} p_t(\omega)^{1-\rho}\phi_t(\omega)L_t(\lambda)w_t(j,\lambda)\frac{\partial \xi_t(\lambda,j)}{\partial p_t(\omega'')} = \sum_{j,\lambda,\kappa} L_t(\lambda)\frac{w_t(j,\lambda)}{\partial p_t(\omega'')}\xi_t(\lambda,j)\pi_t(\lambda,\kappa,\omega,j)$ (summing over all occupation ω, because $\sum_\omega p_t(\omega)^{1-\rho}\phi_t(\omega) = 1$ and $\sum_{\omega,\kappa}\pi_t(\lambda,\kappa,\omega,j) = 1$)

I next show that $\frac{\partial \pi_t(\lambda,\kappa,\omega,j)}{\partial p_t(\omega'')} < 0$, $\frac{\partial w_t(j,\lambda)}{\partial p_t(\omega'')}$ and $\frac{\partial \xi_t(\lambda,j)}{\partial p_t(\omega'')}$ exist and are well defined for all $p_t(\omega) \in \mathbf{R}^n_{++}$. This is the case (using théo I (i)): see how I put it, but I can calculat the derivative for π and w, and use the chain rule for $\frac{\partial \xi_t(\lambda,j)}{\partial p_t(\omega'')} = \frac{\partial \xi_t(\lambda,j)}{\partial p_t(\omega'')}\Big|_{p_t(j)} + \frac{\partial \xi_t(\lambda,j)}{\partial p_t(j)}\frac{\partial p_t(j)}{\partial p_t(\omega'')}$, the last derivative exists for all pomega by theo I (i).

B.7 Derivation of the change in wages over time

The change of the average wage in a group λ in city j is given by, where $x_t(\lambda, \kappa', \omega', j) \equiv \left[T_t(\lambda, \kappa', \omega', j)p_t(\kappa')^{-\frac{\alpha}{1-\alpha}}p_t(\omega')^{\frac{1}{1-\alpha}}\right]^{\theta}$:

$$\hat{w}(\lambda, j) = \frac{w_1}{w_0} = \frac{\left\{\sum_{\kappa',\omega'} x_1(\lambda, \kappa', \omega', j)\right\}^{\frac{1}{\theta}}}{\left\{\sum_{\kappa',\omega'} x_0(\lambda, \kappa', \omega', j)\right\}^{\frac{1}{\theta}}}$$

$$= \frac{\left\{\sum_{\kappa',\omega'} \hat{x}_1(\lambda, \kappa', \omega', j)x_0(\lambda, \kappa', \omega', j)\right\}^{\frac{1}{\theta}}}{\left\{\sum_{\kappa',\omega'} x_0(\lambda, \kappa', \omega', j)\right\}^{\frac{1}{\theta}}}$$

Using equation (2.7):

$$= \frac{\left\{\sum_{\kappa',\omega'} \hat{x}_1(\lambda, \kappa', \omega', j)\pi_0(.)\sum_{\kappa',\omega'} x_0(\lambda, \kappa', \omega', j)\right\}^{\frac{1}{\theta}}}{\left\{\sum_{\kappa',\omega'} x_0(\lambda, \kappa', \omega', j)\right\}^{\frac{1}{\theta}}}$$

$$= \left\{\sum_{\kappa',\omega'} \hat{x}_1(\lambda, \kappa', \omega', j)\pi_0(.)\right\}^{\frac{1}{\theta}}$$

where $\hat{x}_1(\lambda, \kappa', \omega', j) = \left[\hat{T}(\lambda, \kappa', \omega', j)\hat{p}_t(\kappa')^{-\frac{\alpha}{1-\alpha}}\hat{p}_t(\omega')^{\frac{1}{1-\alpha}}\right]^{\theta}$.

B.8 Estimation of the dispersion of idiosyncratic preferences for location with the share of income to housing decreasing with the skill of workers

As I have many skill groups and 3 settlement types, the choice of the reference group is not straightforward. Moreover, a potential bias of the parameter η depends on the choice of the reference group. One source of bias is to impose that the share of income devoted to housing β is independent from the skill group λ. Consider equation (2.9) in time difference in the general case where the share of income devoting to housing depends on the skill group λ, similarly to equation (2.19) it becomes, where to illustrate there are 2 skill groups λ_1 and λ_2 and 2 settlements type j_1 and j_2:

$$\frac{\hat{l}(\lambda_1, j_1)}{\hat{l}(\lambda_2, j_1)}\frac{\hat{l}(\lambda_2, j_2)}{\hat{l}(\lambda_1, j_2)} = \left[\frac{\hat{w}(\lambda_1, j_1)}{\hat{w}(\lambda_2, j_1)}\frac{\hat{w}(\lambda_2, j_2)}{\hat{w}(\lambda_1, j_2)}\right]^{\eta}\left[\frac{\hat{Q}(\lambda_1, j_1)}{\hat{Q}(\lambda_1, j_2)}\frac{\hat{Q}(\lambda_2, j_2)}{\hat{Q}(\lambda_2, j_1)}\frac{\hat{p}(j_2)^{\beta(\lambda_1)}}{\hat{p}(j_1)^{\beta(\lambda_1)}}\frac{\hat{p}(j_1)^{\beta(\lambda_2)}}{\hat{p}(j_2)^{\beta(\lambda_2)}}\right]^{\eta} \tag{B.21}$$

If equation (2.31) is based on equation (B.21) city prices also enter the error term. As equation (2.10) shows, it is positively correlated with wages. Changes in local wages will thus be negatively correlated with the changes in local prices in the error term. Let λ_2 be workers with a university degree and λ_1 workers without it. Let j_1 be large urban centers and j_2, the rest of the country. If the change in housing is larger in urban centers relative to the rest of the country η in equation (2.31) which is the elasticity of labor with respect to wages adjusted for the cost and quality of living will be overestimated. Indeed, if cost of housing increases more in large cities, it means that average wages increases relatively more there. Moreover, if wages increase relatively more for high-skilled workers there, the response of their local population would seem high. However, because they spend a lower share of income for housing, the incidence of housing cost on them is actually lower (which means that their relative increase in real wage is underestimated, and wrongly attributed to η). Under the same condition, if the reference group λ_1 is now workers with a university degree, η is underestimated. Even though the direction of the bias is difficult to characterize in general, its direction depends on the choice of the reference group.

I thus proceed I follow: I compute η in turn with each skill group λ as a reference group. I then take the lowest and the largest estimate and consider it a lower and upper bound.

B.9 Step by step procedure to carry out the counterfactuals

I need to parameterized equation (2.19), 2.23 and 2.26 to perform my counterfactuals. The price of equipment $\frac{\hat{q}(\kappa)^\theta}{\hat{q}(\kappa_1)^\theta}$, change in national population $\hat{L}(\lambda)$, occupation shifter $\hat{a}(\omega)$ and change in the quality of life $\hat{Q}(\lambda, j)$ are exogenous. The relative price of occupation $\frac{\hat{q}(\omega)^\theta}{\hat{q}(\omega_1)^\theta}$ relative local wages $\frac{\hat{w}(\lambda,j)}{\hat{w}(\lambda_1,j)}$ and local population $\frac{\hat{l}(\lambda,j)}{\hat{l}(\lambda,j_1)}\frac{\hat{l}(\lambda_1,j_1)}{\hat{l}(\lambda_1,j)}$ are endogenous. Whereas I borrow the parameter α from the literature, I follow the steps below to complete the numbers I need:

Step 1: directly recovering from the data: initial periods as well as change over time for factor allocation $\pi(\lambda, \kappa, \omega, j)$, wages $w(\lambda, j)$, and population $l(\lambda, j)$, as well as initial labor income in occupation ω: $\zeta_0(\omega) = \sum_{j,\lambda,\kappa} w_t(j, \lambda)L_t(\lambda)\xi_t(\lambda, j)\pi_t(\lambda, \kappa, \omega, j)$.

Step 2: compute $\frac{\hat{q}(\kappa)^\theta}{\hat{q}(\kappa_1)^\theta}$ and $\frac{\hat{q}(\omega)^\theta}{\hat{q}(\omega_1)^\theta}$ according to Section 2.5.2.

Step 3: estimation of η, ρ and θ according to Sections 2.5.3 and 2.5.4.

Step 4: recovering the change in residual labor productivity $\frac{\hat{T}(\lambda)}{\hat{T}(\lambda_1)}$ with equation (2.23), of the change in settlement perceived quality of life $\left[\frac{\hat{Q}(\lambda',j')}{\hat{Q}(\lambda',j)}\frac{\hat{Q}(\lambda',j')}{\hat{Q}(\lambda,j')}\right]^\eta$ with equation (2.19) and occupation shifter $\frac{\hat{a}(\omega)}{\hat{a}(\omega_1)}$ with equation (2.26).

Because those numbers are computed as residuals, I will consider them carefully. Indeed, they will also capture model misspecification. In particular, reasonable changes in the quality of life would be reassuring.

Step 5: computation of the counterfactuals. I take the relative exogenous change of price of computerized equipment $\frac{\hat{q}(\kappa)^\theta}{\hat{q}(\kappa_1)^\theta}$ to be zero between 1979 and the 5 following periods and display the related counterfactual proportion of high-skilled workers in large cities in 1986, 1992, 1999, 2006, 2012, compared with the actual proportion. Also, I display a related counterfactual measure of wage inequality based on equation (2.23) in the 5 period compared to the actual numbers. In this exercise, the exogenous variables are the change in population by skill group $\hat{L}(\lambda)$, the change in occupation shifter $\hat{a}(\omega)$, and the change in perceived quality of life $\hat{Q}(\lambda, j)$. I then use equations (2.19), (2.23) and 2.26 in order to compute the 3 endogenous variables which are the relative price of occupation $\frac{\hat{q}(\omega)^{\theta}}{\hat{q}(\omega_1)^{\theta}}$, relative local wages $\frac{\hat{w}(\lambda,j)}{\hat{w}(\lambda_1,j)}$ and local population $\frac{\hat{l}(\lambda,j)}{\hat{l}(\lambda,j_1)}\frac{\hat{l}(\lambda_1,j_1)}{\hat{l}(\lambda_1,j)}$.

List of Tables

List of Figures

Bibliography

Acemoglu, D. and D. Autor (2011). “Skills, Tasks, and Technologies: Implications for Employment and Earnings”. In: *Handbook of Labor Economics* 4, pp. 1043–1171.

Allen, T. and C. Arkolakis (2014). “Trade and the Topography of the Spatial Economy”. In: *The Quarterly Journal of Economics* 129.3, pp. 1085–1140.

Alvarez, F. and R. E. Jr. Lucas (2007). “General Equilibrium Analysis of the Eaton-Kortum Model of International Trade”. In: *Journal of Monetary Economics* 54, pp. 1726–1768.

Arntz, M., T. Gregory, and U. Zierahn (2016). “The risk of automation for jobs in OECD countries”. In: *OECD iLibrary.*

Autor, D. (2013). “The “Task Approach” to Labor Markets: an Overview”. In: *Journal for Labour Market Research* 46, pp. 185–199.

Bacolod, M., B. S. Blum, and W. C. Strange (2009). “Skills in the City”. In: *Journal of Urban Economics* 65.2, pp. 136–153.

Baum-Snow, N., M. Freedman, and R. Pavan (2018). “Why has Urban Inequality increased”. In: *American Economic Journal Applied Economics* 10.4, pp. 1–42.

Baum-Snow, N. and R. Pavan (2013). “Inequality and City Size”. In: *The Review of Economics and Statistics* 95.5, pp. 1535–1548.

Behrens, K., G. Duranton, and F. Robert-Nicoud (2014). “Productive Cities: Sorting, Selection, and Agglomeration”. In: *Journal of Political Economy* 122.3, pp. 507–553.

Behrens, K. and F. Robert-Nicoud (2015). “Agglomeration Theory with Heterogeneous Agents”. In: *Handbook of Regional and Urban Economics* 5, pp. 171–245.

Berger, T. and C. B. Frey (2016). “Did the Computer Revolution shift the Fortune of U.S. Cities? Technology Shocks and the Geography of New Job”. In: *Regional Science and Urban Economics* 57, pp. 38–45.

Berry, C. R. and E. L. Glaeser (2005). “The Divergence of Human Capital Levels across Cities”. In: *Papers in Regional Science* 84.3, pp. 407–444.

Bond, Steve and Mike Coombes (2007). “2001-based Travel-To-Work Areas Methodology”. In: *Office for National Statistics and Centre for Urban & Regional Development Studies) http://www. statistics. gov. uk/geography/ttwa. asp [08.07. 10].*

Burstein, A., E. Morales, and J. Vogel (2019). “Changes in Between-Group Inequality: Computers, Occupations, and International Trade”. In: *American Economic Journal: Macroeconomics* 11.2, pp. 348–400.

Charlot, S. and G. Duranton (2006). “Cities and Workplace Communication: Some Quantitative French Evidence”. In: *Urban Studies* 43.8, pp. 1365–1394.

Ciccone, A. and G. Peri (2006). “Identifying Human Capital Externalities: Theory with an Application”. In: *Review of Economic Studies* 73.2, pp. 381–412.

Combes, P., G. Duranton, and L. Gobillon (2008). “Spatial Wage Disparities: Sorting Matters!” In: *Journal of Urban Economics* 63.2, pp. 723–742.

Combes, P., G. Duranton, L. Gobillon, D. Puga, and S. Roux (2012). "The productivity advantages of large cities: Distinguishing agglomeration from firm selection". In: *Econometrica* 80.6, pp. 2543–2594.

Combes, P. and L. Gobillon (2015). "The Empirics of Agglomeration Economies". In: *Handbook of Urban and Regional Economics* 5.A, pp. 247–348.

Davis, D. R. and J. Dingel (Mar. 2017). "The Comparative Advantage of Cities".

— (2019). "A Spatial Knowledge Economy". In: *American Economic Review* 109.1, pp. 153–170.

De la Roca, J., G. I. P. Ottaviano, and D. Puga (Apr. 2018). "City of dreams".

De La Roca, J. and D. Puga (2017). "Learning by working in big cities". In: *The Review of Economic Studies* 84.1, pp. 106–142.

Diamond, R. (2016). "The determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000". In: *American Economic Review* 106.3, pp. 479–524.

DiNardo, J. E. and J.-S. Pischke (1997). "The returns to computer use revisited: Have pencils changed the wage structure too?" In: *The Quarterly Journal of Economics* 112.1, pp. 291–303.

Draca, M., R. Sadun, and J. Van Reenen (Aug. 2006). "Productivity and ICT: A Review of the Evidence". In: *CEP Discussion Paper*.

Duranton, G. and D. Puga (2004). "Micro-foundations of urban agglomeration economies". In: *Handbook of regional and urban economics*. Vol. 4, pp. 2063–2117.

Dustmann, C., J. Ludsteck, and U. Schoenberg (2009). "Revisiting the German wage structure". In: *The Quarterly Journal of Economics* 124.2, pp. 843–881.

Eaton, J. and S. Kortum (2002). "Technology, geography, and trade". In: *Econometrica* 70.5, pp. 1741–1779.

Eeckhout, J., R. Pinheiro, and K. Schmidheiny (2014). "Spatial sorting". In: *Journal of Political Economy* 122.3, pp. 554–620.

Elias, P. and M. Birch (1994). "ISCO 88 (COM): A guide for users". In: *Institute for Employment Research, University of Warwick*.

Felstead, A., D. Gallie, F. Green, and H. Inanc (July 2014). *Skills and Employment Surveys Series Dataset, 1986, 1992, 1997, 2001, 2006 and 2012.[data collection]. 2nd Edition. UK Data Service [distributor]. SN: 7467*. URL: http://dx.doi.org/10.5255/UKDA-SN-7467-2.

Felstead, A., D. Gallie, F. Green, and Y. Zhou (2007). *Skills at Work, 1986 to 2006*. ESRC Centre on Skills, Knowledge and Organisational.

Florida, R., C. Mellander, K. Stolarick, and A. Ross (2011). "Cities, skills and wages". In: *Journal of Economic Geography* 12.2, pp. 355–377.

Foray, D. (2004). *Economics of knowledge*. MIT press.

Fujita, M. (1989). "Urban Economic Theory Cambridge University Press". In: *Cambridge, UK*.

Fujita, M. and J. F. Thisse (2013). *Economics of Agglomeration: Cities, Industrial Location, and Globalization*. Second edition. Cambridge, MA.: Cambridge University Press.

Ganong, P. and D. Shoag (2017). "Why has regional income convergence in the US declined?" In: *Journal of Urban Economics* 102, pp. 76–90.

Gaspar, J. and E. L. Glaeser (1998). "Information technology and the future of cities". In: *Journal of urban economics* 43.1, pp. 136–156.

Gaubert, C. (2018). "Firm sorting and agglomeration". In: *American Economic Review* 108.11, pp. 3117–53.

Glaeser, E. L. (2008). *Cities, agglomeration, and spatial equilibrium*. Oxford University Press.

Goos, M., A. Manning, and A. Salomons (2014). "Explaining job polarization: Routine-biased technological change and offshoring". In: *American Economic Review* 104.8, pp. 2509–26.

Gyourko, J., C. Mayer, and T. Sinai (2013). "Superstar cities". In: *American Economic Journal: Economic Policy* 5.4, pp. 167–99.

Hall, A. and B. Beerman (2011). *BIBB/BAuA-Erwerbstätigenbefragung 2006 - Arbeit und Beruf im Wandel. Erwerb und Verwertung beruflicher Qualifikationen. GESIS Datenarchiv, Köln. ZA4820 Datenfile Version 1.0.1, doi:10.4232/1.11072.*

Handbury, J. (2013). "Are poor cities cheap for everyone? non-homotheticity and the cost of living across us cities". In: *The wharton school research paper* 71, pp. 1–054.

Hartmann, J. (2006). *BIBB/BauA Erwerbstätigenbefragung 2005/2006. Strukturkontrolle, Steuerung und Gewichtung der Stichprobe*. München: TNS Infratest.

— (2009). *BIBB-FDZ Daten- und Methodenberichte Nr. 1*. Bonn.

Henderson, J. V. (1974). "The sizes and types of cities". In: *The American Economic Review* 64.4, pp. 640–656.

— (1991). "Urban development: Theory, fact, and illusion". In: *OUP Catalogue*.

Hsieh, C-H., E. Hurst, C. I. Jones, and P.H. Klenow (2019). "The Allocation of Talent and U.S. Economic Growth". In: *University of Chicago, Becker Friedman Institute for Economics Working Paper* 93.

Ioannides, Y. M., H. G. Overman, E. Rossi-Hansberg, and K. Schmidheiny (2008). "The effect of information and communication technologies on urban structure". In: *Economic Policy* 23.54, pp. 202–242.

Karabarbounis, L. and B. Neiman (2013). "The global decline of the labor share". In: *The Quarterly Journal of Economics* 129.1, pp. 61–103.

Kok, S. (2014). "Town and city jobs: How your job is different in another location". In: *Regional Science and Urban Economics* 49, pp. 58–67.

Krugman, Paul (1991). "Increasing returns and economic geography". In: *Journal of political economy* 99.3, pp. 483–499.

Lagakos, D. and M. E. Waugh (2013). "Selection, agriculture, and cross-country productivity differences". In: *American Economic Review* 103.2, pp. 948–80.

Leamer, E. E. and M. Storper (2014). "The economic geography of the internet age". In: *Location of international business activities*. Springer, pp. 63–93.

Lindley, J. and S. Machin (2014). "Spatial changes in labour market inequality". In: *Journal of Urban Economics* 79, pp. 121–138.

Mack, Chris A (2011). "Fifty years of Moore's law". In: *IEEE Transactions on semiconductor manufacturing* 24.2, pp. 202–207.

Mas-Colell, A., M. D. Whinston, J. R. Green, et al. (1995). *Microeconomic theory*. Vol. 1. Oxford university press New York.

Moretti, E. (2013). "Real wage inequality". In: *American Economic Journal: Applied Economics* 5.1, pp. 65–103.

National Statistics, Office for (July 2018). *Annual Survey of Hours and Earnings, 1997-2018: secure access. [data collection]. 13th edition. UK Data Service. SN: 6689*. URL: http://doi.org/10.5255/UKDA-SN-6689-12.

Peri, G. and C. Sparber (2009). "Task specialization, immigration, and wages". In: *American Economic Journal: Applied Economics* 1.3, pp. 135–69.

Polanyi, Michael (2009). *The tacit dimension*. University of Chicago press.

Proost, S. and J. F. Thisse (2019). “What can be learned from spatial economics?” In: *Journal of Economic Literature* 57.3, pp. 575–643.

Redding, S. J. (2016). “Goods trade, factor mobility and welfare”. In: *Journal of International Economics* 101, pp. 148–167.

Redding, S. J. and E. Rossi-Hansberg (2017). “Quantitative spatial economics”. In: *Annual Review of Economics* 9, pp. 21–58.

Rossi-Hansberg, E. (2005). “A spatial theory of trade”. In: *American Economic Review* 95.5, pp. 1464–1491.

Salemink, K., D. Strijker, and G. Bosworth (2017). “Rural development in the digital age: A systematic literature review on unequal ICT availability, adoption, and use in rural areas”. In: *Journal of Rural Studies* 54, pp. 360–371.

Southall, H.R., P. Ell, D.A. Gatley, and I. Gregory (2004). *Great Britain Historical Database : Census Data : Parish-Level Population Statistics, 1801-1951. [data collection]. UK Data Service. SN: 4560*. URL: http://doi.org/10.5255/UKDA-SN-4560-1.

Spanos, Grigorios (2019). “Firm Organization and Productivity Across Locations”. In: *Journal of Urban Economics*.

Spitz-Oener, A. (2006). “Technical change, job tasks, and rising educational demands: Looking outside the wage structure”. In: *Journal of labor economics* 24.2, pp. 235–270.

Thoenig, M. and T. Verdier (2010). “A macroeconomic perspective on Knowledge Management”. In: *Journal of economic growth* 15.1, pp. 33–63.

Tian, L. (2018). “Essays on Spatial Economics”. PhD thesis. Columbia University.

Van Reenen, J., N. Bloom, M. Draca, T. Kretschmer, R. Sadun, H. Overman, and M. Schankerman (2010). “The economic impact of ICT”. In: *Final report*.

Wang, Z. (2016). “Wage growth, ability sorting, and location choice at labor-force entry: New evidence from US Census data”. In: *Journal of Urban Economics* 96, pp. 112–120.

Wheeler, C. H. (2001). “Search, sorting, and urban agglomeration”. In: *Journal of Labor Economics* 19.4, pp. 879–899.

— (2004). “Wage inequality and urban density”. In: *Journal of Economic Geography* 4.4, pp. 421–437.

Williams, R. (2006). “Generalized ordered logit/partial proportional odds models for ordinal dependent variables”. In: *The Stata Journal* 6.1, pp. 58–82.