

Document de travail (Docweb) n° 1807

Tasks, cities and urban wage premia

Anja Grujovic

Novembre 2018

Abstract: Combining rich administrative data for Germany with representative workforce surveys, I find that job task content is robustly predictive of differences in urban wage premia across otherwise observationally equivalent individuals. Based on this, I propose a model where productive advantages of cities are inherently task-specific. Workers of higher ability have a comparative advantage in the tasks whose production benefits the most from urban spillovers. In equilibrium, bigger cities generate larger externalities for more able agents and urban wage premia is skill-biased. I estimate the model using German worker panel data on 336 districts, 331 occupations, 3 education categories and 3 tasks. I find that one standard deviation increase in abstract task intensity is associated with a 5-percentage point increase in the elasticity of earnings with respect to population size. Differences in task-specific urban wage premia remain significant even after controlling for skill premia of larger cities.

Keywords: urban wage premium, agglomeration economies, task production.

Tâches professionnelles, villes et primes salariales urbaines

Abstract : En combinant des données administratives pour l'Allemagne avec des enquêtes représentatives sur la population active, cet article montre que le contenu des tâches d'une profession est un prédicteur robuste des différences de primes salariales urbaines pour des individus partageant les mêmes caractéristiques observables. Partant de ce résultat, cet article propose un modèle dans lequel les avantages productifs des villes sont intrinsèquement spécifiques à une tâche. Les travailleurs les plus aptes ont un avantage comparatif dans les tâches dont la production bénéficie le plus des retombées urbaines. A l'équilibre, les grandes villes génèrent des externalités plus importantes pour les agents les plus aptes, et les primes salariales urbaines sont plus élevées pour les travailleurs plus qualifiés. Le modèle est estimé en utilisant des données longitudinales portant sur des salariés allemands de 336 districts, 331 professions, 3 catégories d'éducation et 3 tâches professionnelles. Nous montrons qu'une augmentation d'un écart-type dans l'intensité d'une tâche abstraite est associée à une augmentation de 5 points de pourcentage de l'élasticité des gains par rapport à la taille de la population. Les différences dans les primes salariales urbaines spécifiques à une tâche restent significatives même en contrôlant pour les primes de compétence des grandes villes.

Mots-clefs : prime salariale urbaine, économies d'agglomération, tâches professionnelles.

¹e-mail: anja.grujovic@cem.es; website: <http://anjagrujovic.com>

²Centro de Estudios Monetarios y Financieros (cemfi), Casado del Alisal 5, 28014 Madrid, Spain

1 Introduction²

Productivity is increasing in agglomeration size. One of the ways this is manifested in the data is in individual earnings. Figure 1 plots average daily earnings for male workers against log population density of West German cities. There is a strong positive relationship between earnings and city size. A ten percent increase in population density corresponds to a one percent increase in average earnings. Moreover, these urban wage premia are heavily skill-biased. In the densest city, Munich, a worker with university degree earns 35% more than an observationally equivalent worker in the least dense district. For a worker with no post-graduate qualifications, this premium is a mere 18%. Previous empirical studies on other developed economies report comparable estimates of urban wage premia and its corresponding skill-bias.³

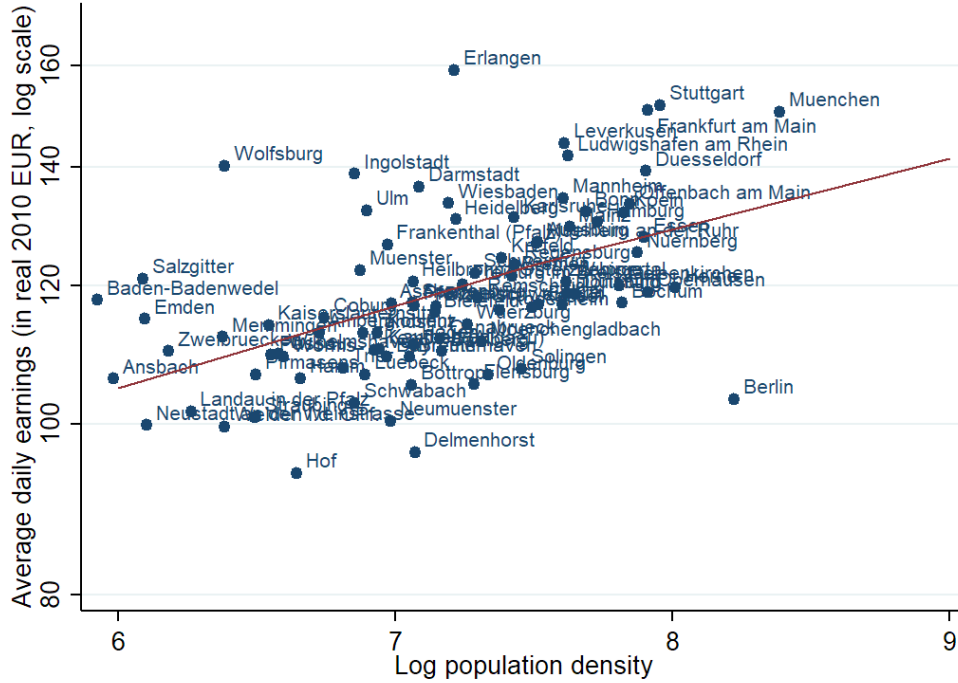
Traditionally, urban economics has focused on estimating skill price differences across locations of different sizes. However, I document that even after controlling for city size differences in skill prices, a significant heterogeneity in urban wage premia remains across occupations. Figure 2 plots such estimates for a set of nine occupations. These estimates are obtained from a pooled OLS regression of individual log wages on a set of observables, as well as on two interaction terms - education interacted with log-population and occupation interacted with log-population. The coefficients on the latter term represent occupation-specific urban wage premia that remain significant even after controlling for any skill price differences across locations of different size. I document that the estimates of these occupation-specific city size premia are strongly correlated with the intensity in abstract task of the occupation. This finding motivates an investigation of how agglomeration economies depend on task demands, and what role these task demands play in explaining the overall skill-premia.

To explain these observations, I build a model where productive advantages of locations are inherently task-specific. I distinguish between three sources of heterogeneity across workers: (i) innate ability, which is given ex-ante and exogenous; (ii) skill, or education,

²I am extremely grateful to Céline Carrère and Frédéric Robert-Nicoud for their guidance and supervision, and to Diego Puga for advice and support. I am also thankful to Manuel Arellano, Giacomo di Giorgi, Joan Monras, Steve Redding, Esteban Rossi-Hansberg, Jose de Sousa, and Lin Tian for their helpful comments, and to numerous seminar and conference participants for their insightful discussions. This research is partially funded by the European Research Council under the European Union's Horizon 2020 Programme (ERC Advanced Grant agreement 695107), by the Swiss National Science Foundation under the Early PostDoc Mobility Fellowship (SNSF Grant number 178064), and by the Spanish Government's "Unit of Excellence Maria de Maeztu" award (Grant number MDM-2016-0684).

³For estimates see Glaeser, 2011, for the United States, Combes, Duranton and Gobillon, 2008, for France, and de la Roca and Puga, 2017, for Spain. Some of their results are discussed later in the text.

Figure 1: Average wages and population density

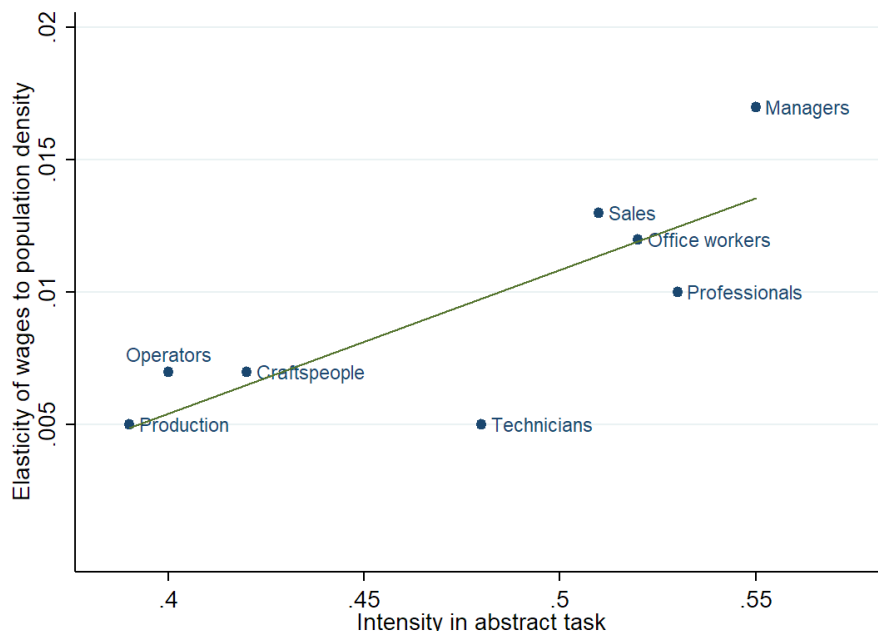


Notes: Average daily wages for West German urban districts over the period 1995-2014, in real 2010 EUR. Sample includes only male workers of Germany nationality between the ages of 20 and 65. Population density is calculated as 2011 population over the surface area of the administrative district.

which is regarded as investment into human capital; and (iii) choice of occupation, where each occupation is regarded as a unique bundle of tasks that must be completed simultaneously for output to be produced. In the model, productive advantages of locations act on task productivities directly. Cities reward the production of tasks differently and these task productivities are allowed to differ across locations due to occupation level indivisibilities. Workers of heterogeneous ability choose an occupation and location in which to produce. Comparative advantage across tasks governs the sorting of heterogeneous individuals into occupations, as in Ricardo-Roy models (Costinot and Vogel, 2015). Frictions to mobility imply imperfect sorting of workers across locations.

In equilibrium, task-specific productivity differences across locations and worker comparative advantage across tasks combine to deliver a rich set of predictions. Workers of higher ability choose to invest more in skill and sort into occupations that are characterised by higher agglomeration economies. As a result, the combination of these two effects gives rise to the

Figure 2: Heterogeneity in urban wage premia by profession, after controlling for city-size differences in returns to education



Notes: Estimated occupation-specific population density wage premia, obtained using a pooled OLS regression of log monthly wages on a set of city, occupation, education, sector, and time indicators, as well as two interaction terms - education interacted with log population density and occupation interacted with log population density. Coefficients on the latter term are represented in the graph. Estimation is based on a sample of full-time working males in Western Germany over the period 1995-2014.

observed skill-bias in urban wage premia. In equilibrium, bigger cities are more skill-intensive and have higher wages.

Using high-quality administrative data for Germany for the period 1995-2014, I find evidence of considerable heterogeneity in city size earning premia across tasks. I find that estimates of task-specific urban wage premia are significant and robust to controlling for skill-level of workers, with largest coefficients on tasks usually considered as more complex. One standard deviation increase in abstract task intensity is associated with a 5 percentage point increase in the elasticity of earnings with respect to population size. These results suggest that variations in job requirements are an important determinant in understanding wage inequalities across space within comparable skill groups.

Existing theoretical frameworks of heterogeneous urban wage premia focus mostly on the differences in the economic value of skill across local labour markets. The concept of

skill in these frameworks, rooted in Becker’s (1964) human capital model, regards skill as a durable investment that can be acquired by the workers through training or education. Though this human capital approach has been successful in capturing the differences in returns to education across different locations, it however remains silent on how and why these geographical differences in skill prices have evolved over time due to changing skill requirements of jobs. My model, building on the task framework approach of Autor, Levy, and Murnane (2003), provides a tractable framework for capturing how skill-biased urban wage premia evolve with varying job requirements.

This paper contributes to the recent theoretical literature studying the distribution of skills and wages across cities.⁴ Building on the task framework, it delivers similar predictions on the spatial variation of skill premia as Davis and Dingle (2012). In turn, the task framework is advantageous as it allows us to study the heterogeneity in urban wage premia across workers of same skill level but different occupations, and, in addition, could be useful for understanding how skill-premia evolve with technological changes to job demands. This paper also complements existing frameworks that model spatial sorting of heterogeneous agents (Behrens, Duranton, Robert-Nicoud, 2014, Davis and Dingle, 2014, amongst others). It also contributes to the vast labour literature that explores shifts in wage structure using the task-based framework.⁵ Unlike my paper, however, existing literature in this field remains silent on how returns to tasks vary across local labour markets.

Empirically, my paper contributes to the growing literature documenting the nature of the relationship between wages, city size and skills.⁶ The estimated magnitudes of the urban wage premium are in line with recent studies that rely on the panel structure of administrative datasets to control for unobservable heterogeneity using individual fixed effects (Glaeser and Maré, 2001; Combes, Duranton, and Gobillon, 2008; and de la Roca and Puga, 2017). My findings are also consistent with Bacold, Blum and Strange’s (2009) evidence that returns to cognitive and people skills are higher in larger agglomerations. They are also in line with Michaels et al. (2016) who find that occupations intensive in abstract tasks are disproportionately present in larger cities.

Applying the task-based approach to the analysis of agglomeration economies could prove relevant for understanding trends in wage inequality. Increasing urbanisation has been identified as a key driver of increasing wage inequality in the U.S. since 1980. As Baum-Snow and Pavan (2013) show, changes in the factor biases of agglomeration economies rationalize at least 80 percent of the more rapid increases in wage inequality in larger cities. In this

⁴For a recent literature review see Duranton and Puga (2004) and Behrens and Robert-Nicoud (2015).

⁵See Autor and Handel (2013) for a recent literature review of the task-based approach.

⁶Rosenthal and Strange (2004) provide a review of the literature.

context, my paper could provide a framework to rationalise and quantify changes in factor biases of agglomeration economies in light of rapid skill-biased technological change.

2 Theoretical framework

This section develops a model in which task-driven productivity differences across locations give rise to skill-biased agglomeration economies. I consider an economy in which a continuum of heterogeneous individuals choose an occupation and a location in which to produce. There is a continuum of available occupations, each defined as a unique bundle of tasks, and a discrete set of locations ($c \in \mathbb{C} = 1, \dots, C$).

2.0.1 Production

Individuals consume a freely traded final good, whose price is set as the numeraire. Producing the final good requires a continuum of occupations indexed by $\sigma \in \Sigma \equiv [\underline{\sigma}, \bar{\sigma}]$. Assuming a CES production function, the output of the final good is given by

$$Q = \left\{ \int_{\sigma \in \Sigma} D(\sigma) Q(\sigma)^{\frac{\epsilon-1}{\epsilon}} d\sigma \right\}^{\frac{\epsilon}{\epsilon-1}}, \quad (1)$$

where $Q(\sigma) \geq 0$ is the quantity of output of occupation σ , which is freely traded, $0 < \epsilon < \infty$ is the elasticity of substitution between occupations, and $D(\sigma)$ is an exogenous demand shifter. Profits by final goods producers are given by

$$\Pi = Q - \int_{\sigma \in \Sigma} p(\sigma) Q(\sigma) d\sigma. \quad (2)$$

Producing occupational output requires only labour that is supplied by a mass of L heterogeneous individuals characterised by their ability α . I denote by $f(\alpha) \geq 0$ the density of workers of skill α and by $\mathbb{A} \equiv [\underline{\alpha}, \bar{\alpha}]$ the set of abilities available in the economy.

I consider an occupation, or a job, to be an indivisible bundle of tasks that must be completed simultaneously for the output to be produced. In this framework, tasks can be thought of as “a unit of work activity that produces output” (Acemoglu and Autor, 2011). I assume that there exists a continuum of tasks, indexed by $\tau \in T \equiv [\underline{\tau}, \bar{\tau}]$. Each occupation requires the input of all tasks, though occupations vary in the intensities at which they demand these tasks. An occupation therefore is defined as a unique bundle of task demands.

The output q of a given occupation σ depends on the intensity at which it uses each tasks, on the ability α of the worker completing the tasks, and on the location c in which it

is being produced:

$$q(\sigma, c; \alpha) = e^{\alpha + \int_{\tau \in T} Z(\tau, c) H(\tau, \sigma) d\tau}, \quad (3)$$

where $Z(\tau, c)$ is the efficiency of task τ production in city c per unit of time, and $H(\tau, \sigma)$ is the share of time an individual in occupation σ is required to spend on task τ . This share of time, or ‘task intensity’, is exogenous, occupation-specific and defined such that $H(\tau, \sigma) \in [0, 1]$ and $\int_{\tau \in T} H(\tau, \sigma) d\tau = 1$ for all σ .

The relative efficiency of task τ production $Z(\tau, c)$, depends only on the location in which it is produced. This parameter is the only source of agglomeration forces in the model, and is taken as given by the worker. There are many reasons why locations might provide productive advantages that are task-specific and I do not make attempts to distinguish between them.⁷ These task efficiencies are allowed to differ across locations due to occupation level indivisibilities.

Importantly, I assume that $Z(\tau, c)$ is twice-differentiable, strictly supermodular in τ and c , and strictly increasing in c . The latter states that cities are indexed by their total factor productivity (TFP), so that higher- c cities are more productive across all tasks. The former dictates comparative advantage in productive advantages of locations, so that the production of higher- τ tasks benefits more from being in higher- c locations than the production of a lower- τ task. As a result, average task efficiency is increasing in τ .

For ease of analysis, I also make assumptions on the functional form of occupation-specific task intensities. I assume that $H(\tau, \sigma)$ is monotonic (non-increasing or non-decreasing), and twice differentiable with $\frac{\partial H(\tau, \sigma)}{\partial \tau \partial \sigma} > 0$. Combined, these assumptions imply that occupations indexed by higher- σ are relatively more intensive in higher- τ tasks. Consequently, the higher- σ occupations are simultaneously more productive on average and have the largest comparative advantage in higher- c locations. For simplicity of exposition, we can then rewrite (3) as

$$q(\sigma, c; \alpha) = e^{\alpha + A(\sigma, c)}, \quad (4)$$

where $A(\sigma, c) = \int_{\tau \in T} Z(\tau, c) H(\tau, \sigma) d\tau$ is the relative occupation-specific productivity across locations. By the assumptions on $Z(\tau, c)$ and $H(\tau, \sigma)$ specified above, $A(\sigma, c)$ is twice-differentiable, strictly supermodular in σ and c , and strictly increasing in c .

Under these assumptions, developing the model in terms of occupation-specific or task-specific agglomeration economies is observationally equivalent. However, formulating this

⁷Several mechanisms can give rise to this effect. Marshall (1920) identifies three key sources of agglomeration economies: labour market pooling, input sharing, and knowledge spillovers. My approach is most closely linked with learning spillovers. One can imagine a setting in which task-specific knowledge is ‘spilt’ or exchanged voluntarily between workers. Such learning spillovers of task-specific human capital would give rise to agglomeration economies which are task-specific.

mechanism on task demands is important for several reasons. Firstly, it allows us to quantify differences in occupation types. Without an observable dimension by which to measure distances between occupations, a definition of an occupation would be entirely meaningless other than as an indicator in an estimation. Secondly, it provides a means to study the evolution of occupation-specific urban wage premia following technological shocks which might alter task demands. This could be particularly relevant in explaining how location specific returns to different occupations have evolved over time following technological shifts in job task requirements over past decades as observed by Autor et al. (2003). Lastly, a mechanism based on tasks is inherently a micro-foundation for occupation-specific agglomeration economies. As such it provides a deeper understanding of factors driving urban wage premia. It also allows for calibration of the model, which can be used to make out-of-sample predictions. In the remainder of paper, I therefore intermittently use both the task definition of agglomeration economies and its occupation-specific equivalent, as is convenient for exposition.

Finally, I define wages. Each individual inelastically supplies one unit of labour, so her income is her physical productivity times the price of the occupational output produced

$$w(\sigma, c; \alpha) = p(\sigma)q(\sigma, c; \alpha). \quad (5)$$

Labour market and preferences

Each individual chooses both an occupation and a location in which to produce. Workers differ in their ability to train for an occupation, and in their idiosyncratic preferences for locations. For ease of exposition, let us consider this process as a sequential decision making, where workers first train for an occupation, before learning their location taste and sorting across cities. This assumption is without loss of generality. The exercise could easily be adapted into a simultaneous decision-making framework, since worker's comparative advantage across occupations does not depend on their choice of location.

In the first instance, individuals choose an occupation to maximise their expected utility, given heterogeneous training costs for learning each task. The total cost of training for an occupation depends on the cost of learning each task and on the intensity at which that task is demanded in the occupation. It is equal to

$$B(\sigma; \alpha) \equiv \int_{\tau \in T} b(\tau, \alpha) H(\tau, \sigma) d\tau \quad (6)$$

where $b(\tau, \alpha) > 0$ is the cost for learning a task τ by an individual of ability α . I assume

that $b(\tau, \alpha)$ is strictly positive, twice differentiable, and strictly submodular in τ and α , and is strictly decreasing in α . This implies that higher- α individuals have a comparative advantage in learning tasks which are indexed with higher- τ - that is, in tasks that benefit more from agglomeration economies (as described in section above). In Appendix A.1, I show that $B(\sigma, \alpha)$ is strictly submodular in σ and α , so that higher α individuals have a comparative advantage in training for higher σ occupations.

Note that this cost $B(\sigma; \alpha)$ can also be seen as an investment into education, which is generally referred to as ‘skill’ in labour economics. As such, this specification allows for an important distinction between skill, a property which can be acquired or invested into, and ability, which is given ex-ante and exogenous. Though in equilibrium they are perfectly correlated, these are two separate concepts. Also note that, as before, it is equivalent whether we had made the assumption of submodularity at task level, $b(\tau, \alpha)$, or directly at occupation level, $B(\sigma, \alpha)$. The benefit of micro-founding the assumption at task level allows us to capture potentially how skill supply in the economy responds to changes in job task demands.

Once individuals have trained for an occupation, and these costs are sunk, serendipity occurs and individuals learn of their preferences for different locations. Workers then choose a location to maximise their utility, given their occupation. To model this location choice, I make use of the conditional logit model, first formulated in the utility maximization context by McFadden (1973). Thus, I assume that the log utility function in location c for an individual of ability α who has chosen occupation σ is:

$$\ln U(C(\sigma, c; \alpha), N(\sigma, c; \alpha)) = \rho + (1 - \beta) \ln C(\sigma, c; \alpha) + \beta \ln N(\sigma, c; \alpha) + \eta(\alpha, c) \quad (7)$$

where $C(\sigma, c; \alpha)$ is the consumption of final good, $N(\sigma, c; \alpha)$ is the consumption of housing, and $\eta(\alpha, c)$ is an i.i.d. preference for a location.⁸ I assume that $\eta(\alpha, c)$ is drawn from a Type I Extreme Value distribution with a shape parameter λ . Utility is maximised subject to budget constraint $C(\sigma, c; \alpha) + r(c)N(\sigma, c; \alpha) \leq w(\sigma, c; \alpha)$, where $r(c)$ is the price of local housing at c .

Given this set-up, the optimisation framework can be specified as follows. First, individuals choose an occupation to maximise their expected utility

$$\max_{\sigma} \mathbb{E}_c[U(\sigma, c; \alpha)] - B(\sigma; \alpha). \quad (8)$$

⁸The i.i.d. preference shocks introduce mobility frictions, which will be relevant for obtaining imperfect sorting of individuals across locations in equilibrium.

Next, they observe their preference shock, and choose a location to maximise their utility

$$\max_c U(c; \alpha, \sigma(\alpha)). \quad (9)$$

Locations and housing market

There is an inelastic supply of housing, N , in each location, owned by absentee landlords. The housing markets are perfectly competitive, so that the market-clearing housing price in location c , consistent with assumption on preferences in (7), is equal to

$$r(c) = \frac{\beta \bar{w}(c) L(c)}{N}. \quad (10)$$

where $\bar{w}(c)$ is the average wage in that location and local population is given by $L(c)$. Note that $L(c)$ also stands for population density, as housing supply is assumed constant across locations.

Definition of a competitive equilibrium

In a competitive equilibrium, individuals maximise their utility, final-good producers and landowners maximise profits, and markets clear.

To qualify the equilibrium, I denote by $g(\sigma, \alpha)$ the share of α -individuals choosing optimally occupation σ , and by $\pi(\sigma, \alpha)$ the share of α -individuals choosing location c . Then, the resulting distribution of individuals across occupations and location is such that

$$g(\sigma, \alpha) > 0 \iff \{\sigma\} \in \arg \max \mathbb{E}_c[U(\sigma, c; \alpha)] - B(\sigma; \alpha) \quad (11)$$

and

$$\pi(c, \alpha) > 0 \iff \{c\} \in \arg \max U(c; \alpha, \sigma(\alpha)). \quad (12)$$

Profit maximisation by final-good producers yields demands for occupational output

$$Q(\sigma) = I \left(\frac{p(\sigma)}{D(\sigma)} \right)^{-\epsilon}, \quad (13)$$

where $I \equiv L \sum_c \int_\sigma \int_\alpha p(\sigma) q(\sigma, c, \alpha) \pi(c, \alpha) g(\sigma, \alpha) f(\alpha) d\alpha d\sigma$ is the total income of all workers in the economy. By free entry, these producer's profits are zero.

Market clearing requires that the housing market clears, that the demand and supply of

occupational output are equal, and that every individual is located somewhere.

$$N = \frac{\beta L}{r(c)} \int_{\sigma \in \Sigma} \int_{\alpha \in A} p(\sigma) q(c, \sigma; \alpha) \pi(c, \alpha) g(\sigma, \alpha) f(\alpha) d\sigma d\alpha \quad \forall c \quad (14)$$

$$Q(\sigma) = \sum_{c \in C} Q(\sigma, c) = L \sum_{c \in C} \int_{\alpha \in A} q(c, \sigma; \alpha) \pi(c, \alpha) g(\sigma, \alpha) f(\alpha) d\alpha \quad \forall \sigma \quad (15)$$

$$f(\alpha) = \sum_{c \in C} \pi(c, \alpha) = \sum_{c \in C} \int_{\sigma \in \Sigma} \pi(c, \alpha) g(\sigma, \alpha) f(\alpha) d\sigma \quad \forall \alpha \quad (16)$$

A competitive equilibrium is a set of functions $Q : \Sigma \rightarrow \mathbb{R}^+$, $g : \Sigma \times \mathbb{A} \rightarrow \mathbb{R}^+$, $\pi : \mathbb{C} \times \mathbb{A} \rightarrow \mathbb{R}^+$, $r : \mathbb{C} \rightarrow \mathbb{R}^+$ and $p : \Sigma \rightarrow \mathbb{R}^+$, such that conditions (10) to (16) hold.

Existence of a competitive equilibrium

To solve for a competitive equilibrium, I start by characterising an individual's optimal choice of occupation and location. Then, I set out conditions for the competitive equilibrium to exist and be unique. Finally, I consider some relevant properties of this equilibrium.

LEMMA 1 (Occupational assignments.) *In a competitive equilibrium, there exists a continuous and strictly increasing matching function $M : \mathbb{A} \rightarrow \Sigma$, such that (i) $g(\sigma, \alpha) > 0$ if and only if $M(\alpha) = \sigma$, and (ii) $M(\underline{\alpha}) = \underline{\sigma}$ and $M(\bar{\alpha}) = \bar{\sigma}$.*

The *proof* of Lemma 1 builds on the analogous exposition in Costinot and Vogel (2010) and is provided in Appendix A.2. Lemma 1 implies that higher- α individuals sort into higher- σ occupations. Given the strict submodularity of $B(\sigma, \alpha)$, it follows immediately that individuals of higher ability have a comparative advantage in occupations with higher training costs. As a consequence, individuals of higher innate ability are overrepresented in occupations which are on average more productive, but that also require higher training costs. Observationally, this would imply that more able individuals are also those with higher levels of skill (where skill is measured by level of education). Note that in what follows, σ is a function of α (as given by the matching function $\sigma = M(\alpha)$ in Lemma 1), but for clarity of exposition I denote it simply as σ .

Given sunk choice of occupation, the workers choose a location to maximise (7). The indirect log utility function for living in each location is

$$\ln v(c, \sigma, \alpha) = \ln p(\sigma) + \alpha + A(\sigma, c) - \beta r(c) + \eta(\alpha, c).$$

I exploit the fact that locational choice does not depend on the price of occupation

nor on the ability of worker (other than through the preference shock $\eta(\alpha, c)$), so that the maximisation problem can be rewritten

$$\arg \max_c \ln p(\sigma) + \alpha + A(\sigma, c) - \beta r(c) + \eta_{\alpha, c} = \arg \max_c \ln V(\sigma, c) + \eta(\alpha, c). \quad (17)$$

where I have defined $\ln V(\sigma, c) \equiv A(\sigma, c) - \beta r(c)$.

By the properties of the conditional logit model, the outcome of this maximisation gives the share of workers of occupation σ that decide to live in city c

$$\pi(\sigma, c) = \frac{V(\sigma, c)^{1/\lambda}}{\sum_k (V(\sigma, k))^{1/\lambda}}. \quad (18)$$

Substituting in for $V(\sigma, c)$, we obtain the following lemma.

LEMMA 2 (Locational assignments) *The share of individuals of occupation σ choosing to produce in city c is*

$$\pi(\sigma, c) = \frac{(e^{A(\sigma, c)} r(c)^{-\beta})^{1/\lambda}}{\sum_k (e^{A(\sigma, k)} r(k)^{-\beta})^{1/\lambda}}.$$

The relative supply of workers of some occupation σ between two cities depends only on the relative (within-occupation) real wages between these locations. Locations with higher average productivity, and lower housing prices, attract more people.

The model being block-recursive, we only need to solve for a vector of housing rents r to determine the full equilibrium.⁹ I proceed as follows. Taking (14), and substituting in using (13), (4) as well as Lemmas 1 and 2, we can obtain the following system of equations that governs housing rents in each location c :

$$r(c) = \left(\sum_k r(k) \right)^{1/\epsilon} \int_{\sigma} \frac{[e^{(1+\lambda)A(\sigma, c)} r(c)^{-\beta}]^{1/\lambda}}{\sum_k [e^{(1+\lambda)A(\sigma, k)} r(k)^{-\beta}]^{1/\lambda}} \left\{ \frac{\sum_k [e^{(1+\lambda)A(\sigma, k)} r(k)^{-\beta}]^{1/\lambda}}{\sum_k [e^{A(\sigma, k)} r(k)^{-\beta}]^{1/\lambda}} \right\}^{1-1/\epsilon} \zeta(\sigma) d\sigma \quad (19)$$

where $\zeta(\sigma) \equiv D(\sigma) \left[\frac{\beta L}{N} e^{M^{-1}(\sigma)} f(M^{-1}(\sigma)) \right]^{1-1/\epsilon}$ is a set of variables and parameters that do not depend on r nor c . The equilibrium vector r then determines $\pi(\sigma, c)$ as per Lemma 2, which in turn is sufficient to characterise all endogenous variables as per (13)-(16).

In what follows, I consider the existence and uniqueness of this equilibrium. I show that

⁹Note that the housing prices $r(c)$ are proportional to the location's total revenue $\bar{w}(c)L(c)$ by (10). Solving for $r(c)$ is equivalent to solving for $\bar{w}(c)L(c)$.

a sufficient condition for the existence and uniqueness of equilibrium is that the elasticity of substitution $\epsilon \geq 1$. In the case of $\epsilon < 1$, when occupations are too strong a complements, it is not clear that a unique equilibrium does exist.

LEMMA 3 (Existence and uniqueness) *For $\epsilon \geq 1$, a competitive equilibrium characterised by equations (10)-(16) exists and is unique.*

Proof. We can consider solving equations in (19) as finding the zeros of an analogous system of “scaffold” functions F . The scaffold function for each location c is obtained by rewriting (19) such that:

$$F(r', r(c)) = \left(\sum_k r'(k) \right)^{\lambda/\epsilon(\lambda+\beta)} \left[\int_{\sigma} e^{(1+\lambda)/\lambda A(\sigma, c)} \frac{\left(\sum_k [e^{(1+\lambda)A(\sigma, k)} r'(k)^{-\beta}]^{1/\lambda} \right)^{-1/\epsilon}}{\left(\sum_k [e^{A(\sigma, k)} r'(k)^{-\beta}]^{1/\lambda} \right)^{1-1/\epsilon}} \zeta(\sigma) d\sigma \right]^{\frac{\lambda}{\lambda+\beta}} - r(c) \quad (20)$$

I verify that the following properties hold for equation (20):

- (i) For all $r' \in \mathbb{R}_{++}^c$, there exists $r(i)$ such that $F(r', r(i)) = 0$,
- (ii) $\frac{\partial F(r', r(i))}{\partial r(i)} \frac{\partial F(r', r(i))}{\partial r(j)} < 0$ for all j ,
- (iii) There exists r' such that for $r(i)$ defined in $F(sr', r(i)) = 0$, $r(i) = o(s)$.

Then, the existence of an equilibrium vector r^* follows from Lemma 1 of Allen, Arkolakis and Li (2015).¹⁰

To prove uniqueness, I check that the following properties also hold:

- (iv) $F(r', r)$ satisfies gross-substitution,
- (v) $F(r', r(i))$ can be decomposed as $F(r', r(i)) = g(r(i)) - h(r(i))$ where $g(r(i))$ and $h(r(i))$ are, respectively, homogeneous of degree α and β , with $\alpha < \beta$.

A sufficient condition for both propositions to hold is $\epsilon \geq 1$.¹¹ Then, uniqueness follows from

¹⁰It is obvious that condition (i) holds since setting $r(c)$ equal to the first term on the right-hand side of the equation implies $F(r', r(i)) = 0$. Condition (ii) also holds because $\frac{\partial F(r', r(i))}{\partial r(i)} = -1$ and $\frac{\partial F(r', r(i))}{\partial r(j)} > 0$. Setting $F(sr', r(i))$ implies $r(i) \propto s^{\frac{\lambda+\epsilon(\lambda+\beta)}{\epsilon(\lambda+\beta)}}$, then condition (iii) also holds. Hence there exists a set of $\{r(i)\}$ that satisfy equation (20).

¹¹That $F(r', r)$ satisfies gross-substitution is evident from the inspection of (20). That property (v) holds as well can be shown as follows. Set $g(r(i))$ equal to the first term of the expression on the right-hand side, and $h(r(i)) = r(i)$. Then $g(r(i))$ is homogeneous of degree $\frac{\lambda}{\epsilon(\lambda+\beta)} - \frac{1}{1-\epsilon}$, and $g(r(i))$ of degree 1. A sufficient condition for $\frac{\lambda}{\epsilon(\lambda+\beta)} - \frac{1}{1-\epsilon} < 1$ is $\epsilon \geq 1$.

Theorem 2 of Allen, Arkolakis and Li (2015). This completes the proof of Lemma 3.

Properties of a competitive equilibrium

I proceed to consider some relevant properties of this equilibrium.¹² First, I examine the relationship between location size $L(c)$ and the location's TFP, indexed by c . Aggregating population shares given in Lemma 2, it can be shown that location size is increasing in c .

LEMMA 4 (Population size) *For a large enough λ/β , population size is increasing in total factor productivity of the location, indexed by c .*

Proof. Combining Lemma 2 and (14), we obtain the following expression for the equilibrium population size of a given location c :

$$L(c) = \left[\underbrace{\int_{\sigma} e^{\frac{1+\lambda}{\lambda} A(\sigma, c) + M^{-1}(\sigma)} p(\sigma) \kappa(\sigma) d\sigma}_{x_1} \right]^{\frac{-\beta}{\lambda+\beta}} \left[\underbrace{\int_{\sigma} e^{\frac{1}{\lambda} A(\sigma, c)} \kappa(\sigma) d\sigma}_{x_2} \right]^{\frac{\lambda-\beta}{\lambda+\beta}},$$

where $\kappa(\sigma) \equiv f(M^{-1}(\sigma)) \left[\sum_k e^{\frac{1}{\lambda} A(\sigma, k)} r(k)^{-\frac{\beta}{\lambda}} \right]^{-1} > 0$ is a collection of occupation-specific variables that are common to all locations.

This function is increasing in c if and only if

$$\frac{\lambda}{\beta} > 1 + \frac{\partial \ln(x_1) / \partial \ln(c)}{\partial \ln(x_2) / \partial \ln(c)}$$

where all terms on the right-hand side are positive by the strict supermodularity assumption on $A(\sigma, c)$.

This inequality can be interpreted as follows. Remember that λ governs dispersion in preferences across locations, and therefore mobility frictions, while β governs the consumption share of housing, which capture congestion costs. The lower is dispersion in preferences (higher λ) and the lower is the share of spending on housing, the more likely it is that agglomeration forces out-power congestion forces and guarantee that location size is increasing in location TFP. If congestion costs are over-bearing however, it is possible to observe locations which are at the top of productivity distribution, but smaller in size, as rents crowd out low-income workers. One such example would be for instance the Silicon Valley, which has some of the highest average wages, but low population density due to high real

¹²Note that, in all of the discussion that follows, none of the results depend on the value of the elasticity of substitution ϵ . Therefore, they hold true for any $0 < \epsilon < \infty$.

estate prices.

Given that the correlation of a location's TFP and its population density is 0.98 for German cities in my sample, I proceed with the assumption that we are in the case where Lemma 4 holds. That is, the city size is increasing in c . Automatically then, for the remainder of this paper, c becomes the index for city's population as well.

LEMMA 5 (Rent schedule) *Rents are increasing in city size, indexed by c .*

Proof. Rearranging (19), the rents in location c can be expressed as

$$r(c) = \left[\int_{\sigma} \tilde{\kappa}(\sigma) e^{\frac{1+\lambda}{\lambda} A(\sigma, c)} d\sigma \right]^{\frac{\lambda}{\lambda+\beta}},$$

where $\tilde{\kappa}(\sigma)$ is again a collection of variables that are independent of c .¹³ By the strict supermodularity on $A(\sigma, c)$, it is evident that $r(c)$ is increasing in c .

Finally we can show that if Lemma 5 holds, average wages are increasing in city size.

LEMMA 6 (A city's productivity) *For a large enough λ , average wages are increasing in city size.*

Proof. By combining Lemmas 4 and 5, we obtain the following expression for average wages in location c :

$$\bar{w}(c) = \left[\int_{\sigma} e^{\frac{1+\lambda}{\lambda} A(\sigma, c) + M^{-1}(\sigma)} p(\sigma) \kappa(\sigma) d\sigma \right] \left[\int_{\sigma} e^{\frac{1}{\lambda} A(\sigma, c)} \kappa(\sigma) d\sigma \right]^{-1}$$

A sufficient condition for this term to be increasing in c is that $\lambda > 0$ be large enough.

Given the primitives of the model, we can now derive two key properties of interest, which will serve as the basis for the empirical estimation in Section 3.

First, note that by (4) and (5), the log earnings of an individual of ability α , choosing optimally to produce in occupation $\sigma = M(\alpha)$ and location c , is given by:

$$\ln(w(\alpha, \sigma, c)) = \ln(p(\sigma)) + \alpha + \int_{\tau \in T} Z(\tau, c) H(\tau, \sigma) d\tau. \quad (21)$$

where, as we have defined before $A(\sigma, c) \equiv \int_{\tau \in T} Z(\tau, c) H(\tau, \sigma) d\tau$. It is useful to keep this equation in mind as it will serve as a reference for the empirical analysis in next section. It also brings us directly to the first proposition.

¹³Specifically, $\tilde{\kappa}(\sigma) \equiv [\sum_k r(k)]^{1/\epsilon} \left\{ \sum_k [e^{(1+\lambda)A(\sigma, k)} r(k)^{-\beta}]^{1/\lambda} \right\}^{-1/\epsilon} \left\{ \sum_k [e^{A(\sigma, k)} r(k)^{-\beta}]^{1/\lambda} \right\}^{1/\epsilon-1} \zeta(\sigma)$.

PROPOSITION 1 (Ability-biased urban wage premia) *Wages are log-supermodular in worker ability α and city size c .*

Proof. By (21), for any $\alpha' > \alpha$ and $c' > c$:

$$\ln \left[\frac{w(\alpha', M(\alpha'), c') \cdot w(\alpha, M(\alpha), c)}{w(\alpha', M(\alpha'), c) \cdot w(\alpha, M(\alpha), c')} \right] = A(M(\alpha'), c') + A(M(\alpha), c) - A(M(\alpha'), c) - A(M(\alpha), c') > 0$$

by strict supermodularity of $A(\sigma, c)$ and by monotonically increasing $M(\alpha)$.

Occupations which benefit the most from agglomeration economies, are also those in which high-ability individuals possess a comparative advantage. In equilibrium, this gives rise to the urban-wage premia that is increasing in worker ex-ante ability. Since high- α individuals are also those that invest more into skill, then *wages are also log-supermodular in skill and city size*.¹⁴

This complementarity in ability and city size also drives the sorting of workers across locations of different size, as next proposition states.

PROPOSITION 2 (Sorting across locations) *Distribution of workers is log-supermodular in worker ability and city size.*

Proof. By Lemma 2, we have that for any two workers $\alpha' > \alpha$ and locations $c' > c$:

$$\ln \left(\frac{\pi(M(\alpha'), c')}{\pi(M(\alpha'), c)} \frac{\pi(M(\alpha), c)}{\pi(M(\alpha), c')} \right) = \frac{1}{\lambda} [A(M(\alpha'), c') + A(M(\alpha), c) - A(M(\alpha'), c) - A(M(\alpha), c')] > 0.$$

This expression is strictly positive by strict supermodularity on $A(\sigma, c)$ and monotonically increasing $M(\alpha)$.

3 Estimation of the model

In the previous sections I showed how task-specific agglomeration economies can give rise to urban wage premia which are skill-biased in equilibrium. In this section, I bring the model to data. Specifically I wish to test (i) if there are significant differences in agglomeration effects across tasks, and if so, (ii) how these task-specific urban productivity advantages relate to the observed skill-bias in urban earnings. Additionally, I test whether these differences in task agglomeration economies can explain sorting patterns of workers across locations, as the model would be predict.

¹⁴Recall that $B(\sigma, \alpha)$ is increasing in σ by assumption. Then $B(M(\alpha), \alpha)$ is increasing in α by Lemma 1.

3.1 Specification: Task specific urban wage premia

I start by considering Proposition I. Direct estimation of equation (21) requires controlling for unobservable worker ability when estimating the relationship between wages and city size. I follow the approach used by Glaeser and Maré (2001), Combes, Duranton, and Gobillon (2008), and de la Roca and Puga (2017) who estimate static urban wage premia while including worker fixed-effects to control for unobserved heterogeneity. In addition to their specification, I also allow for heterogeneity in task-specific city size earning premia. Using a panel dataset of individual earnings, I estimate the following baseline specification:

$$\ln(w_{c\sigma(it)}) = \mu_c + \mu_i + \mu_\sigma + \mu_t + \mathbb{X}_{it}\beta + \sum_{\tau} \delta_{\tau} h_{\tau\sigma(it)} \times \ln(dens_c) + \varepsilon_{it} \quad (22)$$

where $w_{c\sigma(it)}$ is the wage of an individual i at a time t who is observed to be working in location c and occupation σ ; μ_c and μ_σ are indicators for city and occupation, while μ_i and μ_t are individual and time fixed-effects, respectively; \mathbb{X}_{it} is a set of worker time-varying characteristics such as tenure, experience and sector indicators, as well as an indicator for the highest level of education achieved; $h_{\tau\sigma(it)}$ is the observed share of time spent on task τ by an occupation σ ; $dens_c$ is the population density of city c in which the individual is currently working; and ε_{it} is an error term.

Parameter δ_{τ} captures task-specific agglomeration economies. The indicator for city μ_c absorbs the differences in productivity across locations for the reference occupation.¹⁵ The indicator for occupation μ_σ captures the differences in productivity across occupations in the reference location.¹⁶ Therefore the parameter on the interacted task-intensity and location density, δ_{τ} , captures any remaining heterogeneity in city size earnings premia across occupations. The strength of the task-based approach is that we can quantify these differences in agglomeration effects across occupations along a tractable dimension - the task intensity.

A possible source of bias in this estimation is that the task-based approach might be capturing differences to factor prices across locations. If the true mechanism of agglomeration effects is through the skill-bias, the estimates of δ_{τ} would be upward biased for tasks which are intensively used by skilled occupations. To control for this potential confounding effect,

¹⁵The reference occupation is the one with the highest intensity in the dropped task.

¹⁶In addition to controlling for differences in average returns per task in the reference location, the occupation indicators also might capture any additional occupation-specific productive advantages not covered by my model. That is why it is preferable to control for occupation averages rather than task averages.

I run a robustness check and control for the skill-bias in agglomeration effects by adding the term $\delta_s \mathbb{I}_{sit} \times \ln(dens_c)$ to the baseline specification (22), where \mathbb{I}_{sit} is an indicator function for the level of education of worker i at time t . The parameter δ_s then measures variation in city size earning premia across skill groups, not captured by variation in task demands.

3.2 Specification: Sorting across locations

In addition, I wish to test whether the observed local labour force composition is consistent with Proposition II. The model predicts that occupations intensive in high- δ_τ tasks should be over-represented in larger cities. Specifically, by Lemma 2, the relative sorting rate between two occupations $\sigma' > \sigma$ into some city c is

$$\ln \left(\frac{\pi(\sigma', c)}{\pi(\sigma, c)} \right) = \frac{1}{\lambda} \int_{\tau \in T} Z(\tau, c) [H(\tau, \sigma') - H(\tau, \sigma)] d\tau + \ln \frac{\sum_k (e^{A(\sigma, k)} r(k)^{-\beta})^{1/\lambda}}{\sum_k (e^{A(\sigma', k)} r(k)^{-\beta})^{1/\lambda}}. \quad (23)$$

To estimate, I run directly the equivalent of equation (23) on a cross-section of observed differences in sorting between each occupation-pair into some location c :

$$\ln \left(\frac{\pi(\sigma', c)}{\pi(\sigma, c)} \right) = \nu_\sigma + \nu_{\sigma'} + \sum_\tau \theta_\tau (h_{\tau\sigma} - h_{\tau\sigma'}) \times \ln(dens_c) + \iota_{\sigma\sigma'c} \quad (24)$$

where $\pi(\sigma, c)$ is the fraction of individuals of occupation σ working in location c ; ν_σ is the indicator for occupation σ , which controls for the term $\nu_\sigma = \ln \left(\sum_k (e^{A(\sigma, k)} r(k)^{-\beta})^{1/\lambda} \right)$ in equation (23); and $\iota_{\sigma\sigma'c}$ is the error term. The estimated parameter θ_τ governs sorting patterns across tasks. It captures how differences in task intensities between two occupations correlate with their relative sorting rates into larger cities. We should expect the estimates of θ_τ to be correlated with the estimates of δ_τ from (22), since according to the theoretical framework $\theta_\tau = \lambda \delta_\tau = \frac{\partial Z(\tau, c)}{\partial c}$. In other words, an occupation's intensity in the tasks that benefit more from agglomeration effects, should be predictive of its higher sorting into larger cities.

4 Data

To estimate wage differentials across tasks and locations, while controlling for any unobservable individual characteristics, I require a dataset that follows individuals both across time and across locations. The social security records for Germany (SIAB) fulfils these needs. Two additional advantages make the German dataset particularly appropriate. Firstly, the SIAB administrative dataset contains an extremely detailed measure of occupations, coded at

around 331 different values. This makes it particularly amenable to analysing occupation-level differences in earnings at a level which cannot be achieved by alternative social security datasets. Secondly, Germany is one of the few countries with representative data on task content of occupations available through rich labour market surveys - The Qualification and Career Surveys (QCS). These surveys are available in six cross-sections in the period of 1979-2012, and are coded on the same occupation classification as SIAB. By mapping along the dimension of 331 occupation codes, I am able to approximate the task content of workers in the SIAB panel dataset.

4.1 SIAB

The dataset “Sample of Integrated Labour Market Biographies (SIAB)” is provided by the Institute for Employment Research (IAB) of the German Federal Employment Agency (BA). This rolling panel dataset covers a 2% random sample of administrative social security records from 1975 to 2014 and is representative for 80% of the German workforce. It covers all employees subject to social security contributions, which includes all white- and blue-collar workers as well as apprentices, but excludes civil servants, the self-employed and individuals performing military service.

The dataset includes 1,757,925 individuals whose employment histories are covered in 58,220,255 lines of data. The unit of observation in the data is any change in the individual’s employment status, which includes changes in occupation, wage or type of contract within the same firm. As such, the data in this spell-based format is exact to the day. I convert this dataset into a monthly panel that includes each individual’s occupation, wage, and work location as well as some socio-demographics for the longest spell in that month.

Work location of every employment spell is observable at the level of administrative districts. There are 401 administrative districts in Germany of which 107 are “urban” and 294 are “rural” districts.¹⁷ By definition, urban districts are cities that constitute districts in their own right. Rural districts on the other hand may comprise one or more smaller towns. Seeing as some smaller towns are comprised within the rural districts, I keep all districts in the analysis sample. This choice is also motivated by the observation that elasticity of wages to population does not appear to be discontinuous between rural and urban districts, as can be seen in Figure 1. In the appendix, I run robustness tests on the subset of urban districts only.

¹⁷District shapefile can be seen in maps in Appendix B.4, which are discussed later in the paper.

4.2 QCS

To obtain data on the intensity of tasks performed by each occupation, I make use of the BIBB Qualification and Career Survey (QCS) by the German Federal Institute for Vocational Training (Bundesinstitut für Berufsbildung; BIBB) for the year 2006. Amongst other question, survey respondents were required to indicate how relevant a set of 16 general tasks is to their job. Following Spitz-Oener (2006) I group the 16 tasks into three categories: abstract, routine or manual. Appendix B.1 shows the mapping of tasks. By standardising this variable, I obtain an estimate of the average time spent on each of the three task types for 301 occupational groups included in the dataset.

Since both SIAB and QCS are based on same occupation identifiers, I am able to merge the QCS occupation-level averages of task intensities into the SIAB dataset along the occupation dimension. As there might be concerns about individual level variation in task intensities across locations, I run a regression of factors determining the task intensities across individuals in the QCS dataset. The estimates are shown in Appendix B.2. As can be seen, there are significant differences in task intensities across city size, but this significance disappears once we control for occupation (except for some estimates on the manual task). This finding would suggest that most of the heterogeneity in task intensity across locations can be explained by sorting of occupations across cities. As such, averaging task intensities by occupation should not lead to any biases from this perspective.

The resulting panel dataset includes full employment histories, with data on wage growth, occupational and sociodemographic changes, as well as occupation-specific data on task intensities.

4.3 Sample selection

I limit the sample to the period between 1995-2014. This is done for two reasons. Firstly, following unification, data for the entire Berlin district (which is partially included under East Germany) only became credible in 1995. Secondly, as I am using the data on task intensities from the 2006 QCS wave, this represents a neat mid-point in the period studied. Furthermore I restrict the sample under analysis to West Germany only. Due to obvious historical reasons, there are considerable differences in wages between East and West Germany that persist even to this day. Average wages in East Germany are significantly lower than those in the West, even after controlling for city size and observable worker characteristics, as Figure 4 in Appendix B.5 makes evident. For robustness of exercise and comparability to studies on other countries, I restrict my analysis to West Germany only. Finally, I limit the sample to full-time working, males of Germany nationality between the age of 20 and 65.

An apparent particularity of Germany is that its capital and second city by density, Berlin, is estimated to have one of the lowest fixed-effects in the sample. The reason is evidently historical.¹⁸ Despite it being an outlier, I keep Berlin in the sample for the entire analysis as at the very worst it could be introducing a downward bias on the relationship between earnings premia and city size, and so goes against the mechanisms I wish to estimate.

The daily wages, obtained directly from the SIAB dataset, are calculated based on the fixed-period wages reported by the employer and the duration of the (unsplit) original notification period in calendar days. These wages are converted into real daily wages using annual consumer-price indices for West Germany. Additionally, I drop observations where real daily wage is below 10EUR, in 2010 real terms, as such observations are assumed unreasonable as described in Büttner and Rässler (2008). In what follows, when I refer to wages, it is implied that they are in logs and corrected for inflation.

Finally, wages are top-censored at the censoring level which differs by year. I compute the wages above the censoring threshold using the same method as Card, Heining, and Kline (2013), with the exception that I treat districts as they treat firms. Therefore, I run a series of 800 tobit imputations for 4 age groups (20-29, 30-39, 40-49, 50-65), 10 aggregate occupation groups, and 20 years, separately. In each tobit estimation, the observed censored log wage is regressed on a constant, age, education, and an indicator variable for each month. Furthermore, I include the individual's average log wages (excluding the current period) and the fraction of top-censored wage observations over her career (excluding current censoring status). Instead of including firm mean log-wages as in Card, Heining and Kline (2013), I include instead annual mean log-wages in the district as well as the log density of the district population (where density is calculated as population over area).¹⁹ Given the predicted values $X'\beta$ from the tobit regressions, as well as the estimated standard deviation σ , I proceed to impute the censored wages w^s as follows: $w^s = X'\beta + \sigma\Phi^{-1}[\kappa + u(1 - \kappa)]$, where $u \sim U[0, 1]$ and $\kappa = \Phi[(s - X'\beta)/\sigma]$ and s is the official censoring limit. Censored wages are then replaced by these imputed values, while uncensored wages are kept as in the

¹⁸ The Economist covers this phenomenon in a recent article, calling out Berlin for being “poor but sexy.” In the article entitled “Why is Berlin so dysfunctional? - Unlike other capitals, Germany’s is a drain on the rest of the country,” The Economist goes on to explain that this situation was not always the case. Before the second world war, Berlin was a prominent industrial hub of the country. However, following the division, Berlin was no longer an attractive business environment which led most of the firms to relocate their factories to West Germany. Consequently, even after the wall fell, most of the firms, now well established in West Germany, had little incentive to move back. Instead, Berlin attracted artists and bohemians that flocked into abandoned factories and warehouses, turning Berlin into an art hub of Europe. As such, Berlin is unique among major European capitals to make the country poorer on average. Without Berlin, Germany’s GDP per person would be 0.2% higher.

¹⁹ I also include a dummy whenever the individual is observed only once in the entire sample, in which case the mean life-time wage and fraction of censored observations take the value of the sample mean.

original data source. Appendix B.3 lists selected percentiles of the distribution of simulated earnings.²⁰

Due to considerable amount of missing values for education level, I correct these by grouping education into three main categories, as in Fitzernberger, Osikominu, and Völter (2006). The corrected educational categories can take the values: low (without postsecondary education), medium (apprenticeship or Abitur) and high (university degree). I then drop all the observations for which the corrected educational category is missing. This results in substantial attrition, taking the sample from 48,098,327 to 41,703,558 observations.

For certain estimation steps, it is convenient to work with aggregated occupation groups. For this purpose, I create a new variable which aggregates the 3-digit occupations into nine professional groups following the classification in Böhm, von Gaudecker and Schran (2017). The resulting variable, which I refer to as a “profession,” to differentiate it from the 3-digit occupation coding, can take the value of any of the following nine categories: Managers, Professionals, Technicians, Craftspeople, Sale personnel, Office workers, Production workers, Operators/labourers, and Service personnel.

Finally, I drop all observations for which the 3-digit occupation is not available. This is the case for 655,870 observations. I also drop all occupations for which there is no information on task intensities in the QCS dataset, which leads to a loss of an additional 804,257 observations. The final sample, which covers the working histories of German men in all districts of West Germany in the period of 1995-2014 counts 40,243,431 observations. In parts of the analysis where I limit the sample to urban districts only for robustness checks, the sample decreases further to 16,497,568 observations.

4.4 Summary statistics and descriptive patterns

Table 1 provides some descriptive statistics of the dataset. For a clear overview, professions are ordered by increasing intensity of the abstract task, relative to non-abstract tasks. Production profession spends the least amount of time doing abstract tasks - 39% of the time, while Managers perform abstract tasks most frequently - 55% of the time. The opposite holds true for both routine and manual tasks. Several observations can be made just by looking at these descriptives, that go in support of my model’s features.

Firstly, as the model would predict, there is a strong correlation between the intensity of the abstract task and the level of education. The majority of the workers in top-abstract professions (Managers and Professionals) have the highest education degree, the

²⁰Using equivalent extrapolation method on the Spanish social security data, De la Roca and Puga (2017) and De la Roca (2017) show that extrapolated wages match well the observed wages for the years in which observations were not capped.

Table 1: Descriptive statistics, by profession

	Production	Operators	Craftspeople	Service	Technician	Sales	Office workers	Professionals	Managers	Overall
Task intensity										
Abstract	39%	40%	42%	44%	48%	51%	52%	53%	55%	45%
Routine	32%	30%	29%	24%	27%	24%	23%	23%	22%	27%
Manual	29%	30%	29%	32%	25%	25%	25%	24%	23%	27%
Education level										
Low	20%	24%	7%	12%	2%	3%	3%	1%	1%	10%
Medium	78%	73%	90%	72%	78%	77%	71%	32%	40%	69%
High	2%	3%	3%	17%	20%	20%	26%	67%	59%	21%
Average daily wage (in real 2010 EUR)	98	87	97	94	140	125	129	167	185	113
Profession intensity by population density (pop/km^2)										
$\ln(dens) < 7.0$	1.15	1.05	1.15	0.90	0.96	0.93	0.86	0.80	0.79	
$7.0 < \ln(dens) < 7.5$	0.90	0.98	0.84	1.02	1.13	1.09	1.07	1.18	1.05	
$7.5 < \ln(dens) < 7.9$	0.77	0.93	0.75	1.12	1.11	1.14	1.31	1.27	1.30	
$7.9 < \ln(dens)$	0.54	0.85	0.64	1.39	0.97	1.12	1.32	1.62	1.73	
Number of obs. (in millions)	9.80	4.55	6.16	3.03	2.51	2.74	3.42	6.25	1.78	40.24

Note: Table shows averages by profession group over the period 1995-2014 for all West German districts. Professions are ordered by abstract task intensity, ranging from lowest to highest. First part of the table, task intensity, shows the average time spent on each task by profession. Second part, gives the share of workers in each profession by level of education. Profession intensity is calculated as the share of workers of a given profession working in that location group, relative to the share of total population working in said location group ($\frac{\pi(\sigma,c)}{\int_{\sigma} \pi(\sigma,c) d\sigma}$).

majority of those in middle-abstract professions have a medium-level degree, while almost the entire work-force in the bottom three abstract professions (Production, Craftspeople and Operators) only have a low- or medium-education level. The gradient of education level is monotonically increasing across professions, along the dimension of the abstract task. This can be seen more clearly from Table 2. Here, individuals of each education group are ranked according to quartiles of the frequency at which they perform the abstract task. Individuals in lowest education are predominatnly in the lowest abstract task quartile, while the opposite holds true for the high education group. These observations are in line with Lemma 1.

Table 2: Abstract task intensity by education group

	Education		
	Low	Medium	High
Abstract task quartile			
1st	59%	25%	2%
2nd	25%	29%	3%
3rd	12%	28%	31%
4th	4%	18%	64%

Note: Table shows the repartition of workers by education level into occupations by quartiles of task intensity. Averages are calculated for all West German districts, over the period 1995-2014.

Secondly, wages are increasing in the intensity of abstract task, as can be observed from Table 1. Managers benefit from highest average daily wages, while Operators earn the least on average. Bar few exceptions, wages are monotonically increasing between the two extremes of abstract task intensity.

Thirdly, professions more intensive in abstract tasks are over-represented in denser locations. Table 1 displays the profession intensity by location density.²¹ As purported by the model, densest locations are most intensive in high-abstract professions. Cities with log population density above 7.9 (equivalent to 3,000 individuals per squared kilometre), have overrepresented shares of workers in abstract professions such as Service, Sales, Professionals and Managers. Low-density locations, on the other hand, are intensive in non-abstract professions, such as Production and Crafts. For a better geographical overview, I show the geographical repartition of average task intensities by district in Appendix B.4. Here, districts are ranked according to the average time spent on each task by their working population. As can

²¹ Profession intensity is defined as the share of a profession in a given location, relative to total population share in that location ($\frac{\pi(\sigma,c)}{\int_{\sigma} \pi(\sigma,c)}$).

be observed from the maps, denser districts (represented by smaller surface area) tend to specialise in abstract tasks, while the more spread-out districts (larger surface area) specialise predominantly in manual tasks.

5 Empirical Results

Before turning to the results of the baseline estimations, I first run some descriptive regressions to show comparability to previous studies and to highlight the role occupations play in explaining the heterogeneity in city size earning premia. An impatient reader can skip directly to sections 5.2 and 5.3 for main results.

5.1 Occupations and urban wage premia: Descriptive regressions

For illustration purposes, I first run a simple pooled ordinary least squares regression to estimate the average static earning premium per location:

$$\ln(w_{ict}) = \mu_c + \mu_t + X_{it}\beta + \xi_{it}, \quad (25)$$

omitting all other variables from (22). Time-varying individual observables such as experience, tenure and sector indicators are included, but not occupation indicators. Column 1 of Table 3 reports the results. All coefficients are of the expected sign.

Next, as in de la Roca and Puga (2017), I regress the estimates of city indicators from Column 1 on log city size. Results of this regression are shown in Column 2. The elasticity of the earnings premium with respect to city size is estimated to be 0.0373. When I restrict the sample to urban districts only, the estimated elasticity increases to 0.065, as shown in Table 10 in the Appendix B.6. Both of these estimates are within the range of previous results in comparable regressions. Combes, Duranton, Gobillon, and Roux (2010) estimate an elasticity of 0.051 for France, Glaeser and Resseger (2010) find an elasticity of 0.041 for United States, and De la Roca and Puga (2017) obtain an estimate of 0.046 for Spain.

Appendix B.5 show the plots of these estimated city fixed-effects in Column 1 against log city size. In this Appendix, Figure 5 shows the estimated fixed-effects of all districts in the sample, while Figure 6 shows the equivalent estimates when the sample is restricted to urban districts only. There is a strong and positive linear relationship between city size and earnings premia, even after controlling for all observables other than the occupation, and this relationship remains robust regardless of restrictions on district type.

Next, I consider the role of sorting of occupations across locations in explaining city size wage premia. In column 3, I re-estimate (25) while in addition allowing for 331 occupation

Table 3: Descriptive regression: estimation of city-size wage premia

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	Log earnings	City indicator coefficients column (1)	Log earnings	City indicator coefficients column (2)	Log earnings	City indicator coefficients column (3)
Log city density		0.0373*** (0.0027)		0.0290*** (0.0023)		0.0142*** (0.0016)
City indicators	Yes		Yes		Yes	
Occupation indicators			Yes		Yes	
Worker fixed-effects					Yes	
Experience	0.030*** (0.0000)		0.025*** (0.0000)		0.036*** (0.0000)	
Experience ²	-0.001*** (0.0000)		-0.001*** (0.0000)		-0.001*** (0.0000)	
Medium education	0.196*** (0.0002)		0.101*** (0.0002)			
High education	0.560*** (0.0002)		0.281*** (0.0002)			
Observations	41,044,262	326	41,044,262	326	41,044,262	326
Adj. R ²	0.47	0.37	0.57	0.33	0.20	0.20

Notes: All specifications include a constant term. Columns (1), (3), and (5) include month and year indicators, two-digit sector indicators, as well as a linear and a non-linear term for days of tenure. There are 326 city indicators and 331 occupation indicators. Coefficients are reported with robust standard errors in parenthesis, which are clustered by worker in columns (1), (3) and (5). ***, **, and * indicate significance at the 1, 5, and 10 percent levels. The R² reported in column (6) is within workers. Worker values of experience and tenure are calculated on the basis of actual months worked and expressed in years.

indicators. The estimated elasticity of earnings premium with respect to city size decreases by around 20% to 0.0290 (Column 4). Sorting of workers across locations by occupation profiles has a strong predictive power on wage premia of larger cities. Similar magnitude of the effect is estimated when restricting the sample to urban districts only (see Column 4 in Table 10 in Appendix B.6.).

Further allowing for worker-fixed effects (Column 5) decreases the estimate by another 50%. This drop in earnings elasticity following the introduction of worker fixed-effects is comparable to the that in previous studies: Combes, Duranton, Gobillon, and Roux (2010) find a drop of 35% for France, while de la Roca and Puga (2017) find a decline of 50% for Spain in similar exercises. This significant decrease in the earnings premium relative to city size after controlling for worker fixed-effects would suggests the relevance of sorting by workers across locations on unobserved ability, even within very granulated occupation categories. Combined, controlling for sorting by workers on occupations and unobserved ability explains some 60% of the original elasticity of earnings premia in Column 2. Overall, this would suggest that productivity differences across occupations are an important factor in explaining earning differences across space, though unobserved ability of workers within occupations continues to play a major role.

Finally, I explore how the wage premia of larger cities varies across skill- and profession-groups of workers. Starting with a simple pooled ordinary least squares estimation (as in Column 3 of previous example) I additionally allow for a term capturing skill-bias in agglomeration economies ($\delta_s \mathbb{I}_{sit} \times \ln(dens_c)$). Column 1 in Table (4) shows results. In line with previous empirical studies, I find that productivity gains of bigger cities increase in individual's skill level. A worker with university degree earns 35% more in the densest city (Munich) than an observationally equivalent worker in the least dense district. For a worker with vocational training only, this premium drops to 28%, and for someone with a high-school diploma only, this premium is a mere 18%.

In Column 2, I additionally allow for city size earnings premia to vary with 81 occupation codes.²² Allowing for heterogeneity in earnings premia across types of occupations decreases the estimates of skill-bias in agglomeration economies by roughly one half. These estimates of occupation specific city size premia depend positively on the abstract task intensity of the occupation, as my model would predict. In Column 3, I regress the estimated coefficients of occupation-specific agglomeration premia on the average occupational intensity in abstract task. The coefficient is positive (0.006) and highly significant. If instead we allow the density earning premium to vary directly with abstract task intensity, as I do in Column 4,

²²When all 331 occupation indicators are interacted with log city density, majority of estimates are not significant.

Table 4: Heterogeneity in city size earnings premia

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Log earnings	Log earnings	Occupation x pop-density coefficients column (2)	Log earnings	Log earnings
City indicators	Yes	Yes		Yes	Yes
Occupation indicators	Yes	Yes		Yes	Yes
Worker fixed-effects					Yes
Medium education	0.038*** (0.0008)	0.075*** (0.0008)		0.060*** (0.0008)	
High education	0.168*** (0.0010)	0.212*** (0.0012)		0.237*** (0.0010)	
Medium education x log pop-density	0.010*** (0.0001)	0.004*** (0.0001)		0.006*** (0.0001)	0.006*** (0.0002)
High education x log pop-density	0.017*** (0.0001)	0.010*** (0.0000)		0.007*** (0.0002)	0.012*** (0.0003)
Abstract task x log pop-density				0.103*** (0.0007)	
Routine task x log pop-density					
81 occupation indicators x log pop-density		Yes			
Abstract task			0.006*** (0.0003)		
Observations	41,044,262	41,044,262	81	40,240,350	40,240,350
R2	0.57	0.57	0.06	0.57	0.22

Notes: All specifications include a constant term. Columns (1)-(2) and (5)-(7) also include month and year indicators, two-digit sector indicators, as well as non-linear terms for days of experience and tenure. There are 326 city indicators and 331 occupation indicators, unless otherwise indicated. Task intensities are standardised by their respective standard errors. Coefficients are reported with robust standard errors in parenthesis, which are clustered by worker in all columns except (3) and (4). ***, **, and * indicate significance at the 1, 5, and 10 percent levels. The R^2 reported in column (7) is within workers.

we obtain comparable results. Earnings premia of bigger cities increases in the complexity of work performed, and this heterogeneity in task-specific urban premia explains around one half of the observed skill-bias of agglomeration economies in a pooled ordinary least squares estimation.

Ultimately, I allow for worker-fixed effects in Column 5 to control for sorting of workers on unobservable characteristics. The resulting estimates of skill-bias in city size premia are roughly equivalent to the estimates in Columns 2 and 4 in which controls for heterogeneity in earning premia on observable work characteristics were included. Controlling for unobserved worker ability explains approximately half of the pooled ordinary least squares estimates of skill-bias in agglomeration economies from Column 1.

Finally, I present the main results in the next section. In the baseline regression, which is described in detail in Section 3, I combine all of the controls discussed so far to show that differences in task demands across occupations are a significant and robust factor in explaining heterogeneity of urban wage premia.

5.2 Tasks and urban wage premia: Estimating the model

Table 5 shows the main results. Columns 1 and 2 are estimated as per the baseline specification in (22). Columns 3 and 4 additionally include terms for skill-biased agglomeration effects, to control for any confounding effects of skill on task returns. All specifications include indicators for 326 West German districts, 331 occupation indicators, worker fixed-effects as well as time and sector indicators and quadratic terms for tenure and experience.²³

Attention should be made in how we interpret the results in Columns 1 and 2. Consider Column 1 first - here Manual task is the excluded term. Therefore, the reference occupation in this regression is that which is the most intensive in Manual task. In my sample, this occupation is the “Room and Household Cleaner.” Relative to this reference occupation, an increase in intensity in Abstract task by one standard deviation, while keeping Routine task constant, leads to an increase of 7.8 percentage points in the elasticity of wages to population density. Similarly, relative to the reference occupation, an increase of one standard deviation in the Routine task, keeping Abstract task constant, increases the elasticity of wages to population density by 4.6 percentage points.

The specification in Column 2 reports the effect on wage elasticity of increasing the Abstract task intensity relative to all other tasks. In this specification, both Manual and Routine tasks are excluded. The reference occupation is that which is the least intensive

²³Robustness estimates for the subset of urban districts only are shown in Table 12 in Appendix B.6.

Table 5: Estimation of task-specific population-density earnings premia

	(1)	(2)	(3)	(4)
Dependent variable:	Log earnings	Log earnings	Log earnings	Log earnings
City indicators	Yes	Yes	Yes	Yes
Occupation indicators	Yes	Yes	Yes	Yes
Worker fixed-effects	Yes	Yes	Yes	Yes
Medium education x log city density			0.006*** (0.0002)	0.006*** (0.0002)
High education x log city density			0.012*** (0.0003)	0.012*** (0.0003)
Abstract task x log city density	0.078*** (0.0128)	0.052*** (0.0060)	0.052*** (0.0128)	0.030*** (0.0061)
Routine task x log city density	0.046** (0.0193)		0.039*** (0.0193)	
Observations	41,044,262	41,044,262	41,044,262	41,044,262
R ²	0.21	0.21	0.22	0.22

Notes: All specifications include a constant term, month and year indicators, two-digit sector indicators, as well as a linear and a non-linear term for days of tenure and experience. There are 326 city indicators and 331 occupation indicators. Task intensities are standardised by their respective standard errors. Coefficients are reported with robust standard errors in parenthesis, which are clustered by worker. ***, **, and * indicate significance at the 1, 5, and 10 percent levels. The R² is within workers.

in the Abstract task.²⁴ Relative to this reference occupation, an occupation which is one standard deviation more intensive in the Abstract task (relative to all other Non-Abstract tasks) experiences a city size wage elasticity increase of 5.2 percentage points.

These estimated coefficients are all significant and of expected relative magnitudes. Abstract task is the task that most highly benefits from agglomeration economies, followed by the Routine task, while the Manual task benefits the least. Moreover, the relative magnitude of coefficients on Abstract task between Columns 1 and 2 are supportive of the structure of my model. Increasing the frequency of Abstract task relative to Manual task (while keeping Routine task constant) has a higher positive impact on city size earnings premia (7.8 pp), than increasing the frequency of Abstract task relative to all Non-Abstract tasks (5.2 pp).

In Columns 3 and 4, I additionally allow for education specific city size earnings premia to control for a potential source of bias. The estimates of relative task specific agglomeration effects are robust to this specification. The coefficients on interacted task terms remain significant and of the expected relative magnitude, though are smaller in size than in the baseline regression. Differences in skill prices across locations explain only one third of the task-specific urban premia.

To conclude, I find that there is considerable heterogeneity in city size earning premia across tasks, which remains even after controlling for differences in skill prices. These differences in task-specific urban wage premia are significant and robust to different specifications, and imply that variations in job requirements are an important determinant in understanding wage inequalities across space within comparable skill groups.

5.3 Tasks and urban sorting: Estimating the model

Finally, I show that the sorting patterns of workers by occupation are in line with the predictions of my model. Table 6 gives the results of the specification in (24). Occupations that are intensive in the Abstract task are overrepresented in larger cities, while the opposite holds true for occupations intensive in the Manual task. In validation of Proposition 2, I find that differences in urban earnings premia across tasks are predictive of sorting patterns of occupations across locations.

Intensity in abstract task is positively correlated with sorting into bigger cities. A one standard deviation increase in the relative intensity of abstract task between two occupations, increases the elasticity of relative sorting by 12.6 percentage points (Column 1). To interpret these results, consider two occupations: “Machine operator”, the median occupation in terms of abstract task intensity, and “Interior designer”, an occupation that is one standard

²⁴That occupation is “Miners and mineral workers.”

Table 6: Sorting of workers into locations by task intensity

	(1)	(2)
Dependent variable: $\ln\left(\frac{\pi(\sigma',c)}{\pi(\sigma,c)}\right)$	All districts	
Abstract task difference $(h_{A,\sigma'} - h_{A,\sigma}) \times \ln(\text{dens})$	0.126*** (.0002)	
Manual task difference $(h_{M,\sigma'} - h_{M,\sigma}) \times \ln(\text{dens})$		-0.078*** (.0002)
Occupation σ indicators	Yes	Yes
Occupation σ' indicators	Yes	Yes
Observations	23,066,456	23,066,456
Adj. R^2	0.04	0.04

Notes: Regressions are based on equation (24). $\pi(\sigma, c)$ is calculated as the share of workers of a given occupation σ working in location c , averaged over the period 1995-2014. None of the specifications includes a constant. There are 326 locations and 331 occupations. Coefficients are reported with standard errors in parenthesis.

deviation more intensive in abstract task. The estimated coefficient implies that the proportion of interior designers relative to the proportion of machine operators locating in the densest city, Munich, is 40 percent higher than the same ratio in the median city, Gottingen.

The opposite holds true for the Manual task, as is evident from Column 2. Intensity in Manual task is negatively correlated with occupational probability of sorting into cities. Given that the production of the Manual task was estimated to benefit the least from agglomeration economies (see section 5.2), this correspondence in estimates goes in support of Proposition 2.

6 Conclusion

This paper has shown the relevance of considering the nature of job demands for understanding the heterogeneity in urban wage premia. In the model, task-specific productivity differences across locations and worker comparative advantage across tasks combine to give rise to skill-biased urban wage premia in equilibrium. As such, the paper has suggested a novel approach for micro-founding the skill-bias in agglomeration economies. Empirically, the paper has shown that key properties of this equilibrium are supported in the data. Using high quality administrative data, I show that agglomeration economies differ significantly across tasks and are largest for tasks generally considered as more abstract. These differences in task-specific urban wage premia are considerable in magnitude and remain significant even after controlling for differences in returns to education across cities. In this regard, the task-

based approach shows promise on the quest to disentangling the ‘black’ box that are the agglomeration economies.

References

- [1] Acemoglu, Daron, and David H. Autor. 2011. “Skills, tasks and technologies: Implications for employment and earnings.” In *Handbook of labor economics*, vol. 4, ed. Orley Ashenfelter and David Card, 1043–1171. Amsterdam: Elsevier–North Holland.
- [2] Allen, Treb, Costas Arkolakis and Xiangliang Li (2016) “On the Existence and Uniqueness of Trade Equilibria.” Preliminary draft.
- [3] Alvarez, Fernando, and Robert Jr. Lucas. 2007. “General equilibrium analysis of the Eaton-Kortum model of international trade.” *Journal of Monetary Economics* (Elsevier) 54 (6): 1726–1768.
- [4] Antoni, Manfred, Ganzer, Andreas, vom Berge, Philipp (2016): *Sample of Integrated Labour Market Biographies (SIAB) 19752014*. FDZ-Datenreport, 04/2016 (en), Nuremberg.
- [5] Autor, D. H., F. Levy, and R. J. Murnane. 2003. “The Skill Content of Recent Technological Change: An Empirical Exploration.” *The Quarterly Journal of Economics* (Oxford University Press) 118 (4): 1279–1333.
- [6] Autor, David H, and Michael J Handel. 2013. “Putting Tasks to the Test: Human Capital, Job Tasks, and Wages.” *Source Journal of Labor Economics* 31 (2): 59–96.
- [7] Bacolod, Marigee, Bernardo S. Blum, and William Strange. 2009. “Skills in the city.” *Journal of Urban Economics* 65 (2): 136–153.
- [8] Baum-Snow, Nathaniel, and Ronni Pavan. 2013. “Inequality and City Size.” *Review of Economics and Statistics* (The MIT Press) 95 (5): 1535–1548.
- [9] Becker, Gary. 1964. “Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.”
- [10] Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud. 2014. “Productive Cities: Sorting, Selection, and Agglomeration.” *Journal of Political Economy* 122 (3): 507–553.

- [11] Behrens, Kristian and Frederic Robert-Nicoud. 2015. "Agglomeration Theory with Heterogeneous Agents." in Handbook of Regional and Urban Economics, vol. 5, p.171-245, Elsevier.
- [12] Böhm, Michael, Hans-Martin von Gaudecker and Felix Schran. 2017. "The Polarization of Skill Prices in Germany, 1985 - 2010." Working Paper.
- [13] Büttner, Thomas, and Susanne Rässler. 2008. "Multiple Imputation of Right-Censored Wages in the German IAB Employment Sample Considering Heteroscedasticity."
- [14] Card, David, Jorg Heining, and Patrick Kline. 2012. "Workplace Heterogeneity and the Rise of West German Wage Inequality." Cambridge, MA.
- [15] Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2008. "Spatial wage disparities: Sorting matters!" Journal of Urban Economics 63 (2): 723-742.
- [16] Costinot, Arnaud, and Jonathan Vogel. 2010. "Matching and Inequality in the World Economy." Journal of Political Economy 118(4).
- [17] Davis, Donald R. and Jonathan I. Dingel. (2012). "A spatial knowledge economy." NBER Working Paper 18188.
- [18] Davis, Donald R. and Jonathan I. Dingel. (2014). "The comparative advantage of cities." NBER Working Paper No. 20602.
- [19] De la Roca, Jorge, and Diego Puga. 2017. "Learning by Working in Big Cities." Review of Economic Studies 84 (1): 106-142.
- [20] Duranton, Gilles and Diego Puga. 2004. "Micro-foundations of urban agglomeration economies," in: J. V. Henderson and J. F. Thisse (ed.), Handbook of Regional and Urban Economics, edition 1, volume 4, chapter 48, pages 2063-2117 Elsevier.
- [21] Fitzenberger, Bernd, Osikominu, Aderonke and Völter, Robert. 2006. "Imputation Rules to Improve the Education Variable in the IAB Employment Subsample," Schmollers Jahrbuch : Journal of Applied Social Science Studies / Zeitschrift für Wirtschafts- und Sozialwissenschaften, 126, issue 3, p. 405-436.
- [22] Glaeser, Edward L. (2011). "Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier." The Penguin Press.

- [23] Glaeser, Edward, and David Maré. 2001. "Cities and Skills." *Journal of Labor Economics* 19 (2): 316-422.
- [24] Glaeser, Edward L., and Matthew G. Resseger. 2010. "The Complementarity between Cities and Skills." *Journal of Regional Science* 50 (1): 221-244.
- [25] Hall, Anja, Siefer, Anke, Tiemann, Michael (2014): BIBB/BAuA Employment Survey of the Working Population on Qualification and Working Conditions in Germany 2012. suf 3.0; Research Data Center at BIBB (ed.), GESIS Cologne (data access); Bonn: Federal Institute for Vocational Education and Training. doi:10.7803/501.12.1.1.30
- [26] Hug, Jessica. 2017. "City-biased technological change." GSEM Working Paper Series.
- [27] Kok, Suzanne. 2014. "Town and city jobs: How your job is different in another location," *Regional Science and Urban Economics*, 49, 58-67.
- [28] Marshall, Alfred. 1920. "Principles of Economics." London: Macmillan.
- [29] Mas-Colell, Andreu., Michael Dennis. Whinston, and Jerry R. Green. 1995. "Microeconomic theory." Oxford University Press.
- [30] McFadden, D. 1973. "Conditional logit analysis of qualitative choice behavior."
- [31] Small, Kenneth A., and Harvey S. Rosen. 1981. "Applied Welfare Economics with Discrete Choice Models" *Econometrica*. 49(1), pp. 105-130.
- [32] Rosenthal, Stuart S. and William C. Strange. 2004. "Evidence on the nature and sources of agglomeration economies," in: J. V. Henderson and J. F. Thisse (ed.), *Handbook of Regional and Urban Economics*, edition 1, volume 4, chapter 49, pages 2119-2171 Elsevier.
- [33] Spitz-Oener, Alexandra. 2006. "Technical Change, Job Tasks, and Rising Educational Demands: Looking outside the Wage Structure." *Journal of Labor Economics* 24 (2): 235-270.
- [34] Wang, Zhi. 2016. "Wage growth, ability sorting, and location choice at labor-force entry: New evidence from U.S. Census data." *Journal of Urban Economics* 96 (Supplement C): 112 - 120.

A. Theory Appendix

A.1 Proof that $B(\sigma, \alpha)$ is strictly submodular

By definition:

$$B(\sigma; \alpha) \equiv \int_{\tau \in T} b(\tau, \alpha) H(\tau, \sigma) d\tau \quad (26)$$

where, by assumption, $b(\tau, \alpha) > 0$ is twice differentiable, and strictly submodular in τ and α , and is strictly decreasing in α . In addition, $H(\tau, \sigma)$ is assumed to be monotonic and twice differentiable with $\frac{\partial H(\tau, \sigma)}{\partial \tau \sigma} > 0$. Also $\int_{\tau \in T} H(\tau, \sigma) d\tau = 1$ for all σ .

By definition of submodularity, $B(\sigma; \alpha)$ is submodular if and only if, for any $\sigma' > \sigma$ and $\alpha' > \alpha$:

$$\begin{aligned} B(\sigma, \alpha) + B(\sigma', \alpha') &< B(\sigma', \alpha) + B(\sigma, \alpha') \\ \int_{\tau \in T} b(\tau, \alpha) H(\tau, \sigma) d\tau + \int_{\tau \in T} b(\tau, \alpha') H(\tau, \sigma') d\tau &< \int_{\tau \in T} b(\tau, \alpha) H(\tau, \sigma') d\tau + \int_{\tau \in T} b(\tau, \alpha') H(\tau, \sigma) d\tau \\ \int_{\tau \in T} [b(\tau, \alpha) - b(\tau, \alpha')] [H(\tau, \sigma) - H(\tau, \sigma')] d\tau &< 0 \end{aligned} \quad (27)$$

By assumption of monotonicity on $H(\tau, \sigma)$, its second order cross derivative, and by assumption $\int_{\tau \in T} H(\tau, \sigma) d\tau = 1$ for all σ and the mean value theorem, it must be that there is a $\bar{\tau}$ such that $H(\tau, \sigma) - H(\tau, \sigma') \geq 0$ for all $\tau \leq \bar{\tau}$ and $H(\tau, \sigma) - H(\tau, \sigma') \leq 0$ for all $\tau > \bar{\tau}$. Then the inequality in (27) can be rewritten:

$$\int_{\tau_{\min}}^{\bar{\tau}} \underbrace{[b(\tau, \alpha) - b(\tau, \alpha')]}_{>0 \forall \tau} \underbrace{[H(\tau, \sigma) - H(\tau, \sigma')]}_{\geq 0 \forall \tau} d\tau < \int_{\bar{\tau}}^{\tau_{\max}} \underbrace{[b(\tau, \alpha) - b(\tau, \alpha')]}_{>0 \forall \tau} \underbrace{[H(\tau, \sigma') - H(\tau, \sigma)]}_{\geq 0 \forall \tau} d\tau \quad (28)$$

By assumption $\int_{\tau \in T} H(\tau, \sigma) d\tau = 1$ for all σ , it must be that $\int_{\tau_{\min}}^{\bar{\tau}} H(\tau, \sigma) d\tau = 1 - \int_{\bar{\tau}}^{\tau_{\max}} H(\tau, \sigma) d\tau$. This implies that

$$\int_{\tau_{\min}}^{\bar{\tau}} [H(\tau, \sigma) - H(\tau, \sigma')] d\tau = \int_{\bar{\tau}}^{\tau_{\max}} [H(\tau, \sigma') - H(\tau, \sigma)] d\tau.$$

By submodularity of $b(\tau, \alpha)$, it must be that

$$b(\tau, \alpha) - b(\tau, \alpha') < b(\tau', \alpha) - b(\tau', \alpha') \quad \text{for all } \tau' > \tau.$$

Therefore, the left-hand side of (28) must strictly be lower than the right hand side, which implies that the inequality in (27) must strictly hold. This completes the proof that $B(\sigma, \alpha)$ is strictly submodular. The proof for strict supermodularity of $A(\sigma, \alpha)$ is symmetric.

A.2 Proof of Lemma 1

This proof follows closely the analogous proof of Lemma 1 in Costinot and Vogel (2010), with two exceptions. Firstly, in my model the objective function is strictly supermodular, rather than strictly log-supermodular as is their case. Secondly, unlike in Costinot and Vogel (2010), where agents choose a sector to maximise contemporaneous profits, in my model agents choose an occupation that yields the highest expected maximum utility net of training costs, as set out in (11). The relevant expected maximum utility that I will use in this proof can be shown to be the following, given that the error terms are distributed Type I extreme value:

$$\begin{aligned}
\ln V(\sigma; \alpha) &\equiv \mathbb{E}[\max_c V(\sigma, c; \alpha)] \\
&= \lambda \ln \left[\sum_c V(\sigma, c; \alpha)^{1/\lambda} \right] \quad (\text{as shown by Rosen and Small, 1981}) \\
&= \lambda \ln \left[\sum_c \left(\frac{p(\sigma) e^{\alpha + A(\sigma, c)}}{r(c)^\beta} \right)^{1/\lambda} \right] \\
&= \ln(p(\sigma) e^\alpha) + \ln \left[\sum_c \left(\frac{e^{A(\sigma, c)}}{r(c)^\beta} \right)^{1/\lambda} \right] \\
&= \ln \left[p(\sigma) e^\alpha \tilde{A}(\sigma) \right] \quad \text{where} \quad \tilde{A}(\sigma) \equiv \sum_c \left(\frac{e^{A(\sigma, c)}}{r(c)^\beta} \right)^{1/\lambda} \quad (29)
\end{aligned}$$

As in Costinot and Vogel (2010), I proceed with the proof in five steps. Throughout the proof, I denote $\mathbb{A}(\sigma) \equiv \{\alpha \in \mathbb{A} \mid g(\sigma, \alpha) > 0\}$ and $\Sigma(\alpha) \equiv \{\sigma \in \Sigma \mid g(\sigma, \alpha) > 0\}$.

STEP 1. $\mathbb{A} \neq \emptyset$ for all $\sigma \in \Sigma$ and $\Sigma(\alpha) \neq \emptyset$ for all $\alpha \in \mathbb{A}$.

Conditions (15) and (16), together with $f(\alpha) > 0$ for all α , imply that that $\Sigma(\alpha) \neq \emptyset$ for all $\alpha \in \mathbb{A}$. To show that $\mathbb{A}(\sigma) \neq \emptyset$ for all σ , I proceed by contradiction. Suppose that there exists σ' such that $\mathbb{A}(\sigma') = \emptyset$. Since $\Sigma(\alpha) \neq \emptyset$ for all $\alpha \in \mathbb{A}$, we know that there exists σ such that $\mathbb{A}(\sigma) \neq \emptyset$. Therefore, it must be that $Q(\sigma') = 0$ and $Q(\sigma) > 0$. Then, by condition (13), we have that $p(\sigma)/p(\sigma') = 0$. Since there exists $\alpha \in \mathbb{A}(\sigma)$, we know by condition (11) that the following must hold:

$$\begin{aligned}
\ln \left[p(\sigma) e^\alpha \tilde{A}(\sigma) \right] - B(\sigma, \alpha) &\geq \ln \left[p(\sigma') e^\alpha \tilde{A}(\sigma') \right] - B(\sigma', \alpha) \\
\ln \left(\frac{p(\sigma)}{p(\sigma')} \right) &\geq \ln(\tilde{A}(\sigma')) - \ln(\tilde{A}(\sigma)) + B(\sigma, \alpha) - B(\sigma', \alpha),
\end{aligned}$$

which is a contradiction when the limit of $\ln \left(\frac{p(\sigma)}{p(\sigma')} \right)$ tends to $-\infty$.

STEP 2. $\mathbb{A}(\cdot)$ satisfies the following properties: (i) for any $\sigma \in \Sigma$, $\mathbb{A}(\sigma)$ is a non-empty

interval on $[\underline{\alpha}, \bar{\alpha}]$; and (ii) for any $\sigma' > \sigma$, if $\alpha' \in \mathbb{A}(\sigma')$ and $\alpha \in \mathbb{A}(\sigma)$, then $\alpha' \geq \alpha$.

I proceed by demonstrating property i first. Recall that, as shown in step 1, $\mathbb{A}(\sigma)$ is non-empty. To show that $\mathbb{A}(\sigma)$ is an interval, I proceed by contradiction. Suppose that there exists an occupation σ and three workers of abilities $\alpha_1 < \alpha_2 < \alpha_3$ such that $\alpha_1, \alpha_3 \in \mathbb{A}(\sigma)$ but $\alpha_2 \notin \mathbb{A}(\sigma)$. Since $\Sigma(\alpha_2) \neq \emptyset$ by step 1, it must be that there is a $\sigma' \neq \sigma$ such that $\alpha_2 \in \mathbb{A}(\sigma')$. Suppose now that $\sigma > \sigma'$ (the argument for $\sigma < \sigma'$ would be similar). Condition (11) implies that for α_1 to prefer σ :

$$\ln \left[p(\sigma) e^{\alpha_1} \tilde{A}(\sigma) \right] - B(\sigma, \alpha_1) \geq \ln \left[p(\sigma') e^{\alpha_1} \tilde{A}(\sigma') \right] - B(\sigma', \alpha_1),$$

while for α_2 to prefer σ' :

$$\ln \left[p(\sigma') e^{\alpha_2} \tilde{A}(\sigma') \right] - B(\sigma', \alpha_2) \geq \ln \left[p(\sigma) e^{\alpha_2} \tilde{A}(\sigma) \right] - B(\sigma, \alpha_2).$$

Combining these two inequalities, we obtain $B(\sigma, \alpha_1) + B(\sigma', \alpha_2) \leq B(\sigma', \alpha_1) + B(\sigma, \alpha_2)$, which contradicts $B(\sigma, \alpha)$ strictly submodular. Property i follows.

To show property ii, I proceed again by contradiction. Suppose that there exists $\sigma' > \sigma$ and $\alpha' > \alpha$ such that $\alpha' \in \mathbb{A}(\sigma)$ and $\alpha \in \mathbb{A}(\sigma')$. Using condition (11), it follows, in the same manner as before, that $B(\sigma', \alpha) + B(\sigma, \alpha') \leq B(\sigma, \alpha) + B(\sigma', \alpha')$, which contradicts $B(\sigma, \alpha)$ strictly submodular. Property ii follows.

STEP 3. $\mathbb{A}(\sigma)$ is a singleton for all but a countable subset of Σ .

For proof, see Costinot and Vogel's step 3 in the proof of Lemma 1.

STEP 4. $\Sigma(\alpha)$ is a singleton for all but a countable subset of \mathbb{A} .

For proof, see Costinot and Vogel's step 4.

STEP 5. $\mathbb{A}(\sigma)$ is a singleton for all $\sigma \in \Sigma$.

Proof of step 5 is equivalent to that in Costinot and Vogel (2010), except that the relevant inequality by condition (11) is:

$$\ln \left(\frac{p(\sigma)}{p(\sigma')} \right) \geq \ln(\tilde{A}(\sigma')) - \ln(\tilde{A}(\sigma)) + B(\sigma, \alpha) - B(\sigma', \alpha),$$

which is a contradiction when the limit of $\ln \left(\frac{p(\sigma)}{p(\sigma')} \right)$ tends to $-\infty$.

Step 5 implies the existence of a function $H : \Sigma \rightarrow \mathbb{A}$ such that $g(\sigma, \alpha) > 0$ if and only if $H(\sigma) = \alpha$. By step 2's property ii, H must be weakly increasing. Since $\Sigma(\alpha) \neq \emptyset$ for all $\alpha \in \mathbb{A}$ by step 1, H must also be continuous and satisfy $H(\underline{\sigma}) = \underline{\alpha}$ and $H(\bar{\sigma}) = \bar{\alpha}$. Finally, by step 4, H must be strictly increasing. Therefore, there exists a continuous and strictly increasing function $H : \Sigma \rightarrow \mathbb{A}$ such that (i) $g(\sigma, \alpha) > 0$ if and only if $H(\sigma) = \alpha$ and (ii) $H(\underline{\sigma}) = \underline{\alpha}$ and $H(\bar{\sigma}) = \bar{\alpha}$. To conclude the proof of lemma 1, set $M \equiv H^{-1}$. QED.

B. Data Appendix

B.1 Task mapping

Table 7: Classification of BIBB task items into aggregate task groups

Purchasing, procuring, selling Advertising, marketing, public relations Organising, planning and preparing work processes Developing, researching, constructing Training, instructing, teaching, educating Gathering information, investigating, documenting Providing advice and information	Abstract
Manufacturing, producing goods and commodities Measuring, testing, quality control Monitoring, control of machines, plants, technical processes Transporting, storing, shipping	Routine
Repairing, refurbishing Entertaining, accommodating, preparing food Nursing, caring, healing Protecting, guarding, patrolling, directing traffic Cleaning, removing waste, recycling	Manual

Notes: Table shows classification of tasks from BIBB QCS 2006 wave into three aggregate task groups following Spitz-Oener (2006).

B.2 Analysis of factors influencing task intensity

Table 8: Determinants of task intensity

	Abstract		Routine		Manual	
	(1)	(2)	(3)	(4)	(5)	(6)
High School	0.0192 (0.002)***	0.0074 (0.0015)***	-.0048 (0.0015)***	0.0008 (0.0012)	-.0144 (0.0014)***	-.0081 (0.0012)***
Further training	0.0562 (0.0028)***	0.0286 (0.0022)***	-.0188 (0.0021)***	-.0068 (0.0017)***	-.0374 (0.002)***	-.0218 (0.0017)***
University	0.0979 (0.0021)***	0.0361 (0.0018)***	-.0439 (0.0016)***	-.0087 (0.0014)***	-.0540 (0.0015)***	-.0274 (0.0014)***
Age	00 (0.0001)	00 (00)	-.0002 (00)***	-.0002 (00)***	0.0002 (00)***	0.0001 (00)***
Male	0.0195 (0.0011)***	-.0044 (0.001)***	-.0308 (0.0008)***	-.0033 (0.0008)***	0.0114 (0.0008)***	0.0077 (0.0008)***
Local population:						
2.000 - 4.999	0.0031 (0.0027)	0.0001 (0.002)	-.0007 (0.002)	0.0018 (0.0016)	-.0025 (0.002)	-.0020 (0.0016)
5.000 - 19.999	0.0104 (0.0024)***	0.0033 (0.0018)*	-.0049 (0.0018)***	0.0002 (0.0014)	-.0055 (0.0017)***	-.0035 (0.0014)**
20.000 - 49.999	0.0136 (0.0025)***	0.0036 (0.0019)*	-.0059 (0.0019)***	0.0002 (0.0015)	-.0077 (0.0018)***	-.0037 (0.0015)**
50.000 - 99.999	0.0126 (0.0029)***	0.004 (0.0022)*	-.0072 (0.0021)***	-.0014 (0.0017)	-.0054 (0.0021)***	-.0025 (0.0017)
100.000 - 499.999	0.0189 (0.0026)***	0.0037 (0.0019)*	-.0082 (0.0019)***	0.0007 (0.0015)	-.0107 (0.0018)***	-.0043 (0.0015)***
≥ 500.000	0.0237 (0.003)***	0.003 (0.0022)	-.0127 (0.0022)***	-.0003 (0.0018)	-.0110 (0.0021)***	-.0027 (0.0017)
Occup. dummies		Yes		Yes		Yes
Obs.	19,994	19,994	19,994	19,994	19,994	19,994
Adj. R ²	0.21	0.57	0.15	0.48	0.13	0.45

Notes: Table reports regressions based on the 2006 wave of BIBB QCS. In each column, individual level task intensities are regressed on a series of socio-demographics, occupation and local population size. Standard errors are in parenthesis. Each regression includes a constant.

B.3 Selected percentiles of Tobit extrapolated wages

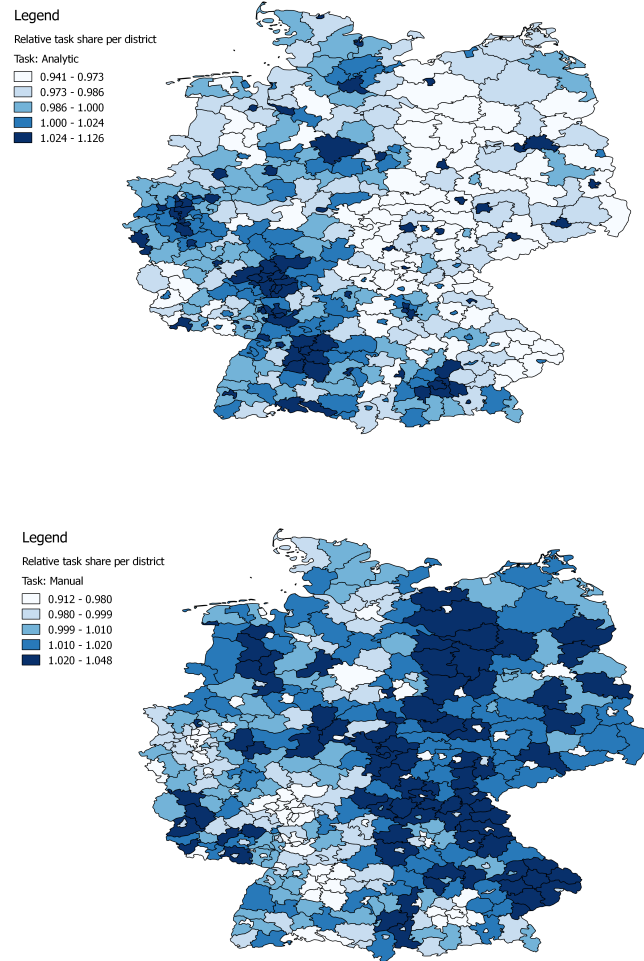
Table 9: Selected percentiles for simulated earnings

	All workers	Skilled workers only
Percentile 1	27%	44%
Percentile 5	47%	70%
Percentile 10	57%	84%
Percentile 25	73%	107%
Percentile 50	92%	138%
Percentile 75	125%	179%
Percentile 90	172%	212%
Percentile 95	199%	234%
Percentile 99	248%	277%

Notes: Monthly earnings are expressed as a percentage of the average wage. Skilled workers are defined as those in the top three professions by abstract task intensity (Managers, Professionals and Office workers).

B.4 Geographical repartition of task intensity

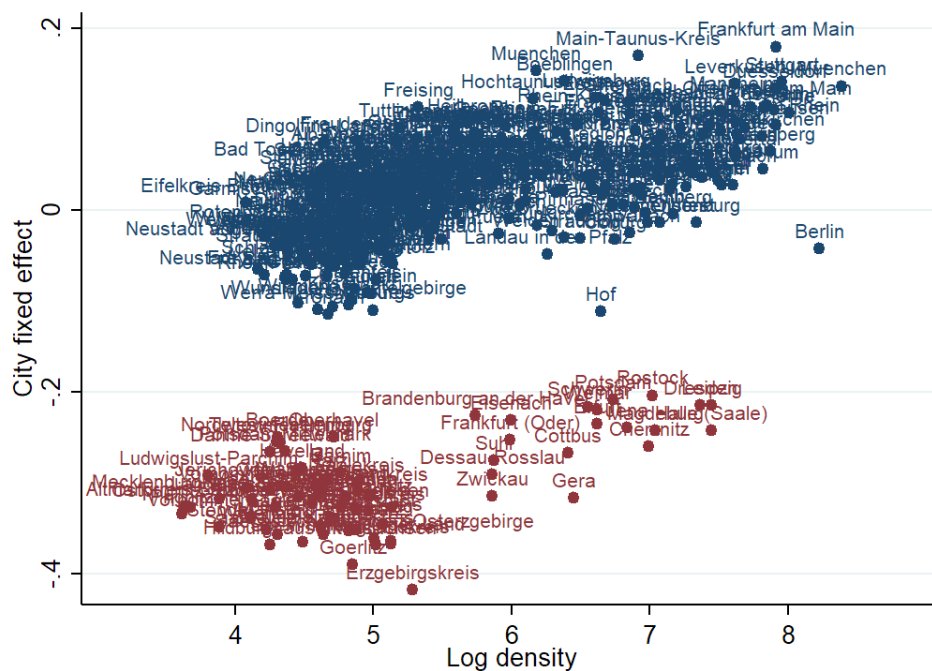
Figure 3: Average task intensity per district: Abstract vs Manual



Notes: Figure shows geographic variation in average task intensity across districts. Task averages are standardised. Values show deviation around national average.

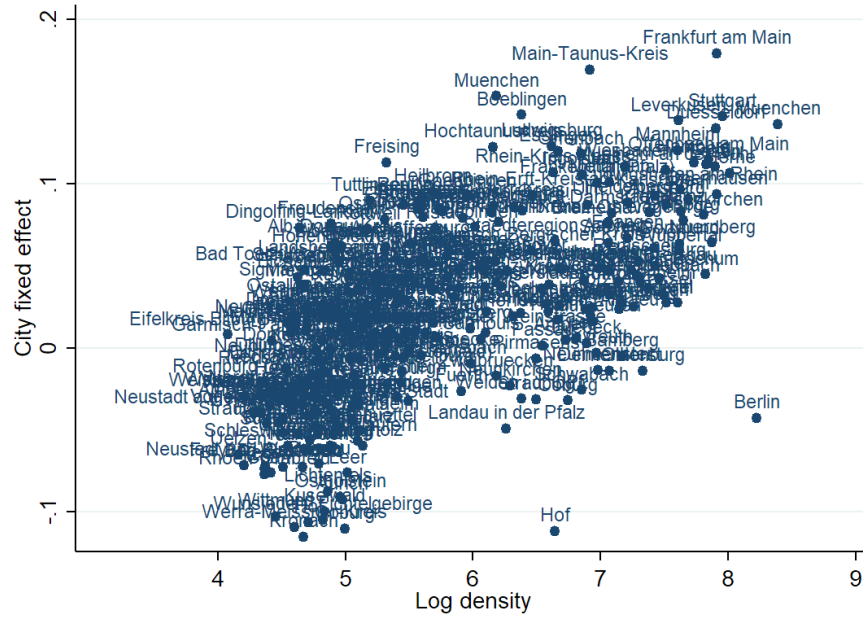
B.5 Estimates of city-fixed effects

Figure 4: City fixed effects, pooled OLS, West and East Germany



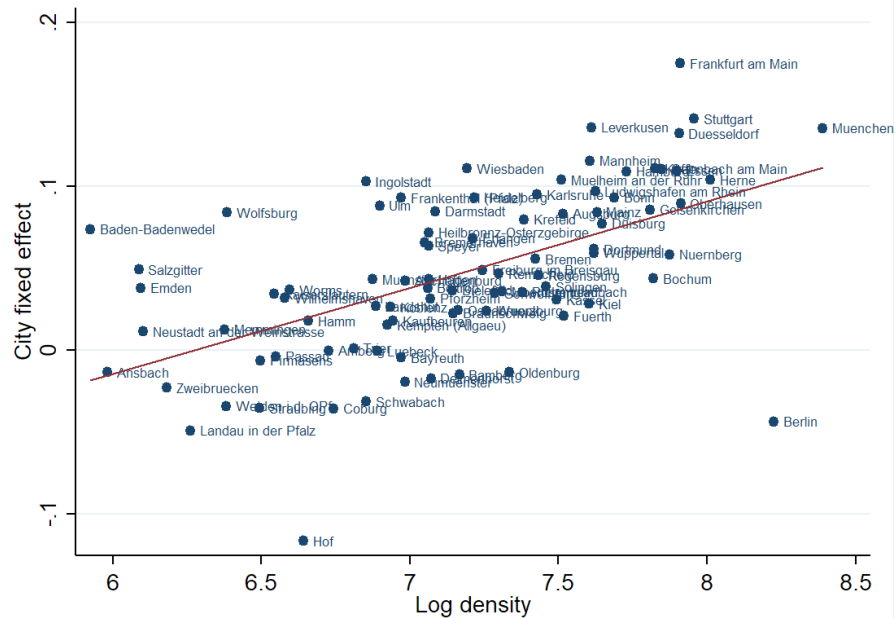
Notes: City fixed effects from the pooled OLS regression in (25) by log population density. West German districts are labelled in blue, East German in red.

Figure 5: City fixed effects, pooled OLS, West Germany only



Notes: City fixed effects from the pooled OLS regression in (25) on West German districts only.

Figure 6: City fixed effects, pooled OLS, West German urban districts only



Notes: City fixed effects from the pooled OLS regression in (25) on West German urban districts only.

Appendix B.6 Robustness estimations on urban districts only

Table 10: Estimation of city-size wage premia, urban districts only

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	Log earnings	City indicator coefficients column (1)	Log earnings	City indicator coefficients column (2)	Log earnings	City indicator coefficients column (3)
Log city density		0.065*** (0.0094)		0.050*** (0.0082)		0.033*** (0.0053)
City indicators	Yes		Yes		Yes	
Occupation indicators			Yes		Yes	
Worker fixed-effects					Yes	
Experience	0.069*** (0.0000)		0.030*** (0.0000)		0.046*** (0.0007)	
Experience ²	-0.001*** (0.0000)		-0.001*** (0.0000)		-0.001*** (0.0000)	
Medium education	0.219*** (0.0003)		0.105*** (0.0002)			
High education	0.628*** (0.0003)		0.292*** (0.0003)			
Observations	16,497,359	89	16,497,359	89	16,497,359	89
Adj. R ²	0.47	0.34	0.58	0.30	0.22	0.30

Notes: All regressions are run on the subset of data for urban districts only. All specifications include a constant term. Columns (1), (3), and (5) include month and year indicators, two-digit sector indicators, as well as a linear and a non-linear term for days of tenure. There are 89 city indicators and 331 occupation indicators. Coefficients are reported with robust standard errors in parenthesis, which are clustered by worker in columns (1), (3) and (5). ***, **, and * indicate significance at the 1, 5, and 10 percent levels. The R² reported in column (6) is within workers. Worker values of experience and tenure are calculated on the basis of actual months worked and expressed in years.

Table 11: Heterogeneity in population-density earnings premia, urban districts only

	(1)	(3)	(4)
Dependent variable:	Log earnings	Log earnings	Log earnings
City indicators	Yes	Yes	Yes
Occupation indicators	Yes	Yes	Yes
Worker fixed-effects			Yes
Medium education	0.090*** (0.0038)	0.158*** (0.0038)	
High education	0.228*** (0.0042)	0.425*** (0.0047)	
Medium education x log city density	0.002*** (0.0005)	-0.007*** (0.0005)	0.006*** (0.0004)
High education x log city density	0.008*** (0.0006)	-0.018*** (0.0006)	0.010*** (0.0005)
Abstract task x log city density		0.242*** (0.0026)	
Routine task x log city density			
Observations	16,497,359	16,497,359	16,497,359
Adj. R ²	0.58	0.58	0.22

Notes: All specifications include a constant term, month and year indicators, two-digit sector indicators, as well as non-linear terms for days of experience and tenure. There are 89 city indicators and 331 occupation indicators. Task intensities are standardised by their respective standard errors. Coefficients are reported with robust standard errors in parenthesis, which are clustered by worker. ***, **, and * indicate significance at the 1, 5, and 10 percent levels. The R² reported in column (4) is within workers.

Table 12: Estimation of task-specific pop.-density earnings premia, urban districts only

	(1)	(2)	(3)	(4)
Dependent variable:	Log earnings	Log earnings	Log earnings	Log earnings
City indicators	Yes	Yes	Yes	Yes
Occupation indicators	Yes	Yes	Yes	Yes
Worker fixed-effects	Yes	Yes	Yes	Yes
Medium education x log city density			0.006*** (0.0004)	0.006*** (0.0004)
High education x log city density			0.010*** (0.0005)	0.010*** (0.0005)
Abstract task x log city density	0.206*** (0.0497)	0.125*** (0.0240)	0.187*** (0.0497)	0.108*** (0.0239)
Routine task x log city density	0.142** (0.0749)		0.138*** (0.0748)	
Observations	16,497,359	16,497,359	16,497,359	16,497,359
R2	0.22	0.22	0.22	0.22

Notes: All specifications include a constant term, month and year indicators, two-digit sector indicators, as well as a linear and a non-linear term for days of tenure. There are 89 city indicators and 331 occupation indicators. Task intensities are standardised by their respective standard errors. Coefficients are reported with robust standard errors in parenthesis, which are clustered by worker. ***, **, and * indicate significance at the 1, 5, and 10 percent levels. The R^2 is within workers.