The Spatial Distribution of Tasks, Skills, and Wages*

Anja Grujovic-Vischer[†] Maximilian Perl[‡]

May 2024

Abstract

Since the 1980s, the segregation of workers by skill across cities has consistently increased with high-skill workers sorting into larger cities. We show that changes in job task content can partially explain the evolution of spatial sorting by worker skill. Building on stylized facts of worker sorting, we propose a framework in which spatial sorting by worker skill arises from more productive cities having a comparative advantage in skill-intensive tasks. In equilibrium, initially more productive locations grow larger and attract more high-skill workers, and the urban wage premium is larger for both high-skilled workers and occupations with a higher share of skill-intensive tasks. Our framework yields two novel predictions. First, the observed skill-bias in the urban wage premium stems from a task-bias in agglomeration economies. Second, a worker's productivity in a specific city is determined by her occupation's task content and the city's comparative advantage in the production of skill-intensive tasks. Consequently, our model shows that the evolution of spatial sorting by skill can be explained by changes in job task content. Using administrative data from Germany, we find support for our model's predictions. The urban wage premium ranges from 1.7 to 2.5 percentage points depending on job task content. Moreover, we observe strong spatial sorting of occupations by their analytical task content.

Keywords: cities, agglomeration, tasks, skills, wages, Germany

JEL Codes: R10, J31, R23

---

*Anja Grujovic-Vischer is extremely grateful to Céline Carrère and Frédéric Robert-Nicoud for their guidance and supervision, and to Diego Puga for advice and support. Maximilian Perl is grateful to Tobias Seidel and Jens Wrona for their excellent supervision. A previous draft – written solely by Anja Grujovic-Vischer – was circulated as a working paper (Grujovic, 2018). Anja Grujovic acknowledges support 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).

[†]Geneva School of Economics and Management (GSEM), University of Geneva. Email: anja.grujovic@gmail.com.

[‡]Economics Department, University of Duisburg-Essen and RTG 2484 Regional Disparities and Economic Policy. E-mail: maximilian.perl@uni-due.de.

1 Introduction

Over the last four decades, skill segregation across cities has increased markedly, with college graduates increasingly sorting into larger cities while workers without college degrees have become concentrated in smaller ones (e.g., Moretti, 2013).[1] This has important implications for welfare inequality between skill groups, as prices, access to amenities, and employment opportunities vary systematically across space (Diamond, 2016; Diamond and Gaubert, 2022). A growing literature attributes this pattern to a widening gap in the urban wage premium between high- and low-skill workers, driven by skill-biased technical change and globalization (e.g., Autor, 2020), further amplified by endogenous amenities (Diamond, 2016).[2]

However, recent evidence suggests that the divergence in the urban wage premium is driven less by worker skill per se than by differences in occupational task content (e.g., Koster and Ozgen, 2021). Consequently, sorting by job tasks has intensified: occupations intensive in analytical and interactive tasks have increasingly concentrated in larger cities (e.g., Bacolod et al., 2009; Michaels et al., 2018). This suggests that changes in the spatial allocation of job tasks may be a central, yet underexplored, mechanism behind observed skill sorting. In this paper, we develop a theoretical framework in which cities differentially reward tasks, workers sort into occupations accordingly, and skill segregation across cities emerges endogenously.

We build a Ricardo–Roy model in which a continuum of workers sorts across both occupations and locations, and show that changes in job task content can account for the evolving patterns of spatial sorting.[3] Our framework embeds the task-based approach of Autor et al. (2003) into a spatial equilibrium setting à la Costinot and Vogel (2010) and Davis and Dingel (2020). We model occupations as bundles of tasks that must be performed to produce output. In equilibrium, workers choose occupations according to their comparative advantage in acquiring the skills required for different task bundles. This generates a natural link between workers' skills and occupational choice, allowing sorting by occupation and sorting by skill to be interpreted as two sides of the

[1]For reviews of the literature on worker sorting by skill and its implications, see Behrens and Robert-Nicoud (2015), Diamond and Gaubert (2022) and Diamond and Suárez Serrato (2025).

[2]At the same time, rising housing costs in larger cities have disproportionately affected less educated and lower-income workers, further reinforcing spatial sorting across cities (e.g., Ganong and Shoag, 2017; Parkhomenko, 2025; Gaubert and Robert-Nicoud, 2025).

[3]For a review of Ricardo–Roy models, see Costinot and Vogel (2015).

same coin. At the same time, workers sort across cities based on locations' comparative advantage in performing different tasks. As a result, a worker's productivity in a given city depends jointly on the task content of her occupation and the task-specific productivity of that location.

This framework provides a novel explanation for the evolution of spatial sorting by skill. When the task content of occupations changes over time, for instance due to task-biased technological change, the geographic distribution of workers across cities adjusts endogenously. Because an occupation's productivity in a given location depends on both its task composition and the location's comparative advantage across tasks, changes in occupational task content translate into changes in spatial sorting patterns across skill groups.

Using administrative employee data for Germany, we begin by documenting two empirical facts on the evolution of spatial sorting. First, sorting by both worker skill and occupational task content increased substantially between 1985 and 2010. Over this period, the local share of college-educated workers and of workers employed in analytical occupations rose more rapidly in larger cities, closely mirroring patterns documented for the United States (e.g., Moretti, 2013; Michaels et al., 2018). Second, we document that the local wage premium for analytical occupations rises with the local share of workers in analytical occupations, which is consistent with (albeit not definitive proof of) a comparative advantage of larger cities in analytical occupations.[4]

This evidence motivates our decision to study spatial sorting through the lens of a Ricardo–Roy framework, in which comparative advantage governs the allocation of workers across both occupations and locations. We model a continuum of heterogeneous workers who sort into occupations based on their comparative advantage in acquiring skills, and across locations based on locations' comparative advantage in the performance of different tasks.

We define occupations as bundles of tasks that must be performed simultaneously by a worker to produce occupation-specific output. To enter an occupation, workers must learn how to perform the associated task bundle by investing in their human capital. Tasks (and thus occupations) differ in their educational requirements, or skill intensity, with more productive tasks on average requiring greater educational investment. Workers, in turn, differ in their ex ante ability, which determines their comparative advantage in learning task bundles: more able workers have a comparative

[4]Similarly, we find that the college wage premium increases in the local share of college graduates.

advantage in acquiring the skills required for more skill-intensive occupations. As a result, the sorting of heterogeneous workers into occupations is governed by their comparative advantage in training costs. Note that we explicitly distinguish between exogenous worker ability and endogenous worker skill (i.e., education), the latter representing an investment decision.

Worker sorting across locations is governed by locations' comparative advantages across tasks. While final occupational output is freely traded, task productivities may differ across individuals, occupations, and locations due to occupation-level indivisibilities. An occupation's productivity is thus determined by its task bundle's composition and a city's productivity in the task contained in its task bundle. These differences in occupation productivities drive the sorting of workers across locations. Workers' idiosyncratic preferences over locations imply that spatial sorting is imperfect.

In equilibrium, these forces jointly produce a spatial economy in which higher-ability workers sort into more skill-intensive occupations, which are concentrated in more productive (and thus larger) cities. The urban wage premium is skill-biased: average wages rise with city size, but productivity gains are greater for higher skill workers. Consequently, bigger cities are also more skill-intensive. Although all workers positively sort into bigger cities, the sorting is stronger for those with greater ability and skill. Because occupational choice is tightly linked to human capital investment, the model also predicts a task-biased urban wage premium and stronger spatial sorting for occupations with more skill-intensive task bundles.

We test the predictions of our model using a rich worker panel covering a 2% sample of German employees, which we merge with detailed data on occupational task content. We characterize an occupation's task bundle using its analytical and manual task intensities, defined as the share of time a typical worker in the occupation spends on analytical and manual tasks, respectively. In the model, occupations differ in the skill intensity of their task bundles; empirically, we capture this dimension using an occupation's analytical task intensity. This measure is both tractable and empirically powerful, capturing the main variation in occupational task content while allowing for parsimonious specifications.[5]

Our empirical results strongly support the model's predictions. After controlling for worker and

[5]The correlation between analytical task intensity and the first principal component of the task space exceeds 0.95.

job characteristics, we find that doubling a city's population is associated with a 1.8% increase in local wages, consistent with prior estimates from other countries. Crucially, and in line with our theoretical framework, the urban wage premium is strongly task- and skill-biased. The wage elasticity with respect to local population ranges from 1.7% for the least analytical (i.e., most manual) occupations to 2.5% for the most analytical occupations, and increases by 0.18 percentage points for a one–standard deviation increase in an occupation's analytical task intensity. These results persist within narrowly defined skill groups and remain robust to controls for skill bias in the urban wage premium.

This paper complements existing theories of spatial sorting that seek to explain why larger cities are more skilled and why the urban wage premium is skill-biased (e.g., Behrens et al., 2014; Eeckhout et al., 2014; Davis and Dingel, 2019, 2020; Gaubert and Robert-Nicoud, 2025).[6] Our framework builds on the idea that locations have a comparative advantage across tasks (and therefore across occupations), and is most closely related to the model in Davis and Dingel (2020). In their framework, higher-skill workers have a comparative advantage in more productive sectors. Agglomeration economies are skill-biased, such that a city's productivity increases with both its population size and average skill level. These forces jointly lead to more skilled workers and more productive sectors sorting into larger cities.[7]

Like Davis and Dingel (2020), we draw on a long tradition of factor-driven comparative advantage models from the international trade literature (e.g., Costinot and Vogel, 2010). However, in contrast to much of the existing urban literature–where productivity advantages act on skill (e.g., Eeckhout et al., 2014; Davis and Dingel, 2019, 2020)–our model features productivity advantages across tasks, with sorting by skill emerging as a by-product of workers sorting into occupations. While our model yields similar predictions for the spatial distribution of skills and skill premia, it offers new predictions for the distribution of tasks and task prices across space. In particular, it allows for differences in an occupation's productivity over time to stem either from variation in a location's comparative advantage or from changes in the occupation's task bundle, providing a

[6] For reviews of these frameworks, see Behrens and Robert-Nicoud (2015) and Diamond and Suárez Serrato (2025).

[7] Specifically, the additional productivity gains offered by larger cities to more skilled workers imply that wages (and therefore the willingness to pay for residing in larger citie) increase with worker skill.

novel explanation for the evolution of spatial sorting: changes in occupational task content.

Moreover, our modeling approach explicitly distinguishes between exogenous ability and education–i.e., "skill"–which is determined endogenously through occupational choice.[8] This allows the model to account for heterogeneity within skill groups, an empirical pattern documented by, e.g., Koster and Ozgen (2021), which previous models have struggled to reproduce and highlights that sorting by occupation and sorting by skill are two sides of the same coin.

This paper also contributes to a growing body of empirical work examining the role of spatial sorting for inequality across and within cities (e.g., Moretti, 2013; Baum-Snow and Pavan, 2013; Diamond, 2016; Baum-Snow et al., 2018; Eckert et al., 2022; Giannone, 2022; Cerina et al., 2022).[9] A core finding in this literature is that changes in the skill bias of agglomeration economies have driven sorting by skill, thereby contributing to spatial wage inequality (e.g., Baum-Snow et al., 2018).[10] Our framework complements this result by showing that the productivity advantages of larger cities for high-skill workers can be decomposed into advantages for specific tasks and job task content. In our model, task-biased structural change would alter cities' productivity advantages across skill groups by changing the task content of jobs–providing a new mechanism for the evolution of wage inequality both across and within cities.

Finally, we contribute to the empirical literature quantifying agglomeration economies (e.g., Ciccone and Hall, 1996; Glaeser and Maré, 2001; Combes et al., 2008; de la Roca and Puga, 2017).[11] In line with the literature on spatial sorting by skill, the empirical literature on agglomeration economies has emphasized the role of skill–generally finding that higher skill workers benefit more from working in larger cities–, but largely omitted the role of occupations.

Only a small number of studies in this literature have studied the role occupations play in determining the urban wage premium (e.g., Gould, 2007; Baum-Snow and Pavan, 2011; Koster and Ozgen, 2021; Perl, 2023). Using a large employee panel and detailed task data, we provide new empirical evidence for Germany that a worker's urban wage premium crucially depends on her

[8] Our definition of skill aligns with the human capital literature following Becker (1964).

[9] For a review, see Diamond and Suárez Serrato (2025).

[10] An alternative explanation relates to the non-homotheticity of housing consumption (Ganong and Shoag, 2017; Parkhomenko, 2025). However, this explanation still requires that higher-skilled workers benefit more from working in larger cities.

[11] For comprehensive reviews, see Rosenthal and Strange (2004) and Combes and Gobillon (2015).

occupation's task content. In Germany, the urban wage premium increases with its analytical task intensity, is absent for the most manual occupations, and varies across tasks within skill groups. Our theoretical framework offers one explanation for the empirical findings of this literature.

The remainder of the paper is organized as follows. Section 2 describes the data and presents stylized facts motivating our modeling choices. Section 3 develops the model. In Section 4, we test our model's predictions empirically. Section 5 concludes. All proofs are presented in the Appendix.

2 Stylized Facts of Worker Sorting

2.1 Data

Employees' Employment Biographies Our principal data source is the *Sample of Integrated Employment Biographies* (SIAB).[12] The SIAB is constructed from social security records and records on recipients of unemployment benefits. It covers the entire employment biographies for a 2% random sample of workers in Germany between 1975 and 2019, excluding only the self-employed and some civil servants exempt from social security contributions. For each job spell, the SIAB reports wages, location, and worker characteristics. We transform the spell data to an annual panel.[13]

Because wages in our data are top-coded at the social security contribution limit, we impute censored values adopting the method of Dauth and Eppelsheimer (2021).[14] We use the covariates from our baseline specification, along with leave-one-out average wages for each worker, local labor market and occupation to predict top-coded wages.

The SIAB reports worker location as the district in which their employing establishment is located. We aggregate the roughly 400 districts to 140 local labor markets using the definition of Kosfeld and Werner (2012). Their definition is comparable to that of Metropolitan Statistical Areas by the US Census Bureau, ensuring that most workers' location corresponds to their residence and

[12] Access to the SIAB is provided by the Institute for Employment Research (IAB). For a detailed description of the version used, see Frodermann (2021).

[13] German social security records are based on annual employer notifications to health insurers. Therefore, wage data refers to average daily wages for the year between notification and there is no advantage of using a higher frequency panel.

[14] In 2019, the contribution limit was 86,000 Euros in West Germany and 80,000 Euros in East Germany.

workplace. Throughout, we use the terms location and city to refer to local labor markets. We complement the SIAB with population and land area statistics from the German statistical office (DeStatis) for the period 1985–2010.

We restrict the sample to workers employed full-time in West Germany between 1985 and 2010, who were aged 16 to 65 and born in 1969 or later (see Appendix B.1 for details on our sample selection). Our final sample comprises over 1.5 million worker-year observations across 108 West German local labor markets. We assess the robustness of our results to these selection criteria in our sensitivity analyses.

Occupational Task Content Following Acemoglu and Autor (2011), we conceptualize occupations as unique bundles of tasks that differ in their skill intensity, specifically in the difficulty of learning how to perform the tasks associated with each occupation. While our model remains agnostic about which tasks are skill-intensive, that is, which tasks link worker skill to cities' productivity advantages, we must make a specific choice for our empirical implementation.

Motivated by Koster and Ozgen (2021), who find that agglomeration economies disproportionately benefit workers in more analytical occupations, and by Acemoglu and Autor (2011), who point out that more analytical occupations are more skill-intensive, we proxy each occupation's task bundle using its analytical task intensity. We draw on data from the employment survey conducted by the Federal Institute for Vocational Education and Training (BIBB; German: Bundesinstitut für Berufsbildung) in 2006.[15] These data report, among other information, each worker's occupation as well as responses to questions on how often they perform 15 tasks on the job, such as using a computer. Following Adda and Dustmann (2023), we aggregate the 15 tasks into two groups, analytical and manual, and define each three-digit occupation group's task intensity as the share of all tasks performed that are classified as analytical. This measure increases with an occupation's skill requirements (see Table C.6) and remains one-dimensional, which facilitates interpretation and implementation. Appendix Table C.5 reports the mapping of tasks to task groups.

[15]The employment surveys of the Federal Institute for Vocational Education and Training (BIBB) are also known as the BIBB/IAB Qualification and Career Surveys.

Table 1: Worker Characteristics by Current Location

	Small Cities	**Medium Cities**	**Large Cities**
Daily Wages (2015 Euros)	122.40	130.37	149.30
	(83.89)	(94.01)	(117.36)
Age (in years)	40.74	41.01	41.21
	(10.16)	(9.89)	(9.56)
Experience (in years)	16.37	16.21	15.68
	(10.57)	(10.32)	(10.00)
Tenure (in years)	9.78	8.65	7.59
	(8.56)	(8.58)	(7.78)
Share Vocational Degree	78.69%	72.58%	63.71%
	(40.95)	(44.61)	(48.08)
Share College Degree	13.92%	18.94%	26.96%
	(34.61)	(39.18)	(44.38)
No. Observations	453,726	777,951	576,516

Notes: Mean and standard deviation of selected variables for our analysis sample drawn from the Sample of Integrated Employment Biographies (SIAB). The observations are worker-year pairs from 1985 to 2010 for male German nationals born after 1969 who are employed full-time in West Germany and have never resided in East Germany. City size groups are defined based on 1985 population figures, using the following cutoffs: fewer than 500,000; between 500,000 and 1,500,000; and more than 1,500,000 inhabitants. Standard deviations are shown in parentheses.

Summary Statistics Table 1 presents summary statistics for our main sample separately for three city size groups (less than 500,000; between 500,000 and 1,500,000; and more than 1,500,000 inhabitants). Workers in larger cities earn higher wages, are more likely to work in analytical jobs, are more likely to have a college degree, and have lower average tenure than workers in smaller cities. However, they are comparable in terms of age and experience.

2.2 The Evolution of Spatial Sorting

Before developing our framework of spatial sorting in the subsequent section, we document a set of stylized facts that motivate our modeling choices.

College graduates increasingly sort into bigger cities. Figure 1 documents the evolution of spatial sorting by worker skill and occupational task content. Panel (a) plots the local share of employees with a college degree against city population in 1985, 1990, 2000, and 2010. For ease of exposition and to satisfy data protection requirements, we divide cities into 20 equal-sized bins based on their population in 1985.[16] Since 1985, the share of college-educated workers in West Germany has increased substantially, rising by 17 percentage points from 3% in 1985 to 20% in 2010.[17] This aggregate trend, however, masks substantial heterogeneity across cities. Over the same period, the share of college graduates in the largest cities increased by 24 percentage points, from 4% to 28%, whereas in the smallest cities it rose by only 7 percentage points, from 1% to 8%. As a result, the semi-elasticity of the local share of college graduates with respect to population increased fivefold between 1985 and 2010, from 0.008 to 0.047. That is, doubling a city's population was associated with a 0.8% increase in the share of college graduates in 1985, compared to a 4.7% increase in 2010.

Panel (b) shows that the evolution of spatial sorting by skill correlates with a comparable evolution in sorting by job task content. Workers in analytical occupations were always more likely to be located in larger cities: in 1985, 40% of workers in the largest cities were employed in predominantly analytical occupations, compared to 29% in the smallest cities. Over time, this pattern intensified. By 2010, nearly 60% of workers in the largest cities were in analytical occupations, compared to 36% in the smallest cities. Occupations are classified as analytical (or manual) based on their task content in 1986–the first year of the BIBB survey–ensuring that changes in occupational classification do not mechanically drive the observed evolution of sorting patterns.

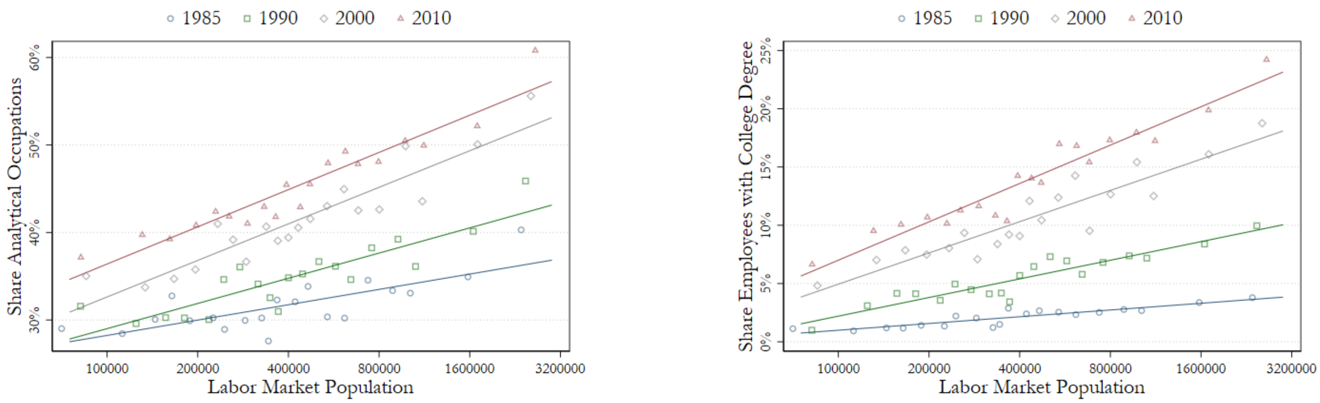
A key force driving spatial sorting is heterogeneity in urban wage premia across skill and occupation groups (e.g., Berry and Glaeser, 2005). Figure 2 plots local log annual wages against log city population separately for college- and non-college-educated workers, as well as for analytical and manual occupations, in 1985, 1990, 2000, and 2010. In all years and for all groups, wages increase with city size, but the strength of this relationship varies substantially across groups and over time. Since at least 1990, the urban wage premium has been consistently larger for analytical occupations than for manual ones, and for college-educated workers than for non-college-educated workers. Moreover, these differences have widened over time. While in 1985 there was little

[16] The IAB requires that reported statistics be calculated from at least 20 individuals. For some small commuting zones in the 1980s, the SIAB covers fewer than 20 full-time workers per year.

[17] The relatively low share of college-educated workers in Germany, compared to other high-income countries, reflects institutional features of the German education system. Occupations that typically require college education elsewhere (e.g., nursing) are instead accessed through vocational training in Germany.

Figure 1: The Spatial Distribution Tasks and Skill

(a) Task-biased Spatial Sorting (b) Skill-biased Spatial Sorting



Notes: Sorting by worker skill and job task content across West German cities, 1985–2010. Panel (a) plots the average share of college-educated workers against log city population for 20 equal-sized bins of commuting zones in 1986, 1990, 2000, and 2010. Panel (b) plots the average share of workers in analytical occupations against log city population for the same bins and years. Bin averages are weighted by commuting-zone employment. Population corresponds to the plotted year, while cities are assigned to bins based on their 1985 population. Occupations are classified as analytical if their analytical task intensity in 1986 exceeds 0.5. The figure is constructed using population data from the German Statistical Office (DeStatis) and the Sample of Integrated Employment Biographies. For details on sample construction, see the text. To ensure the presence of high-skill workers in 1985, the sample used for this figure imposes no restriction on workers' year of birth.

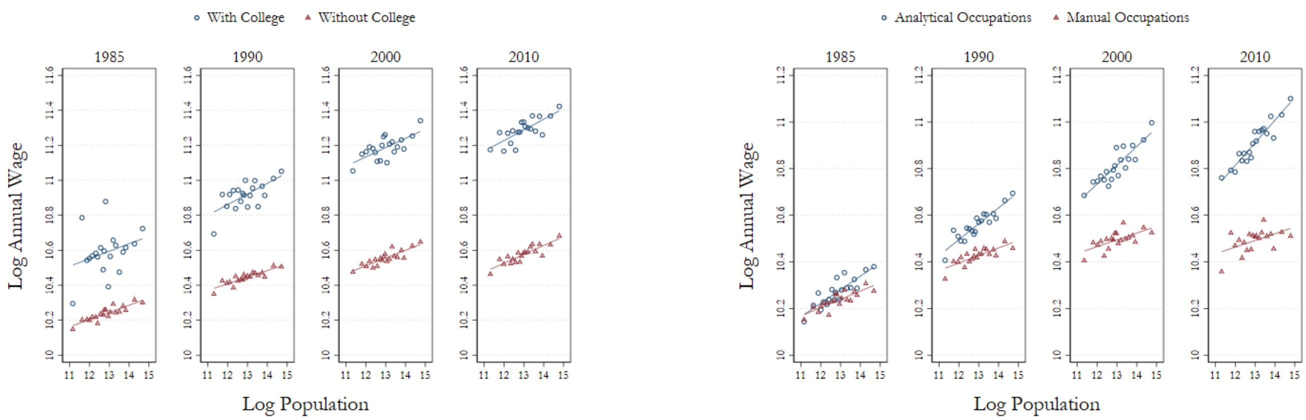
meaningful difference in the urban wage premium across education groups, the task-based difference was already pronounced at that time.[18]

Cities with a larger share of analytical occupations pay relatively higher wages for analytical occupations. In Figure 3, we combine these observations and plot the log ratio of wages in analytical occupations to wages in manual occupations against the local share of workers in analytical occupations for 2010. Each marker corresponds to a local labor market and each marker's size represents the local population. Figure 3 shows that wages in analytical occupations are relatively higher in cities where many employees work in analytical occupations, and these locations are also larger. This is consistent with locations specializing in the production of tasks in which they have a comparative advantage. Motivated by this observation, we develop a framework in which task-specific productive advantages drive spatial sorting in the next section.

[18]Because differences across education groups are less visually salient in Figure 2, Appendix Figure C.5 plots the difference in log mean wages—by commuting-zone population quantiles—between college- and non-college-educated workers.

Figure 2: The Spatial Distribution Skill and Task Prices

(a) Skill-bias in the Urban Wage Premium (b) Task-bias in the Urban Wage Premium



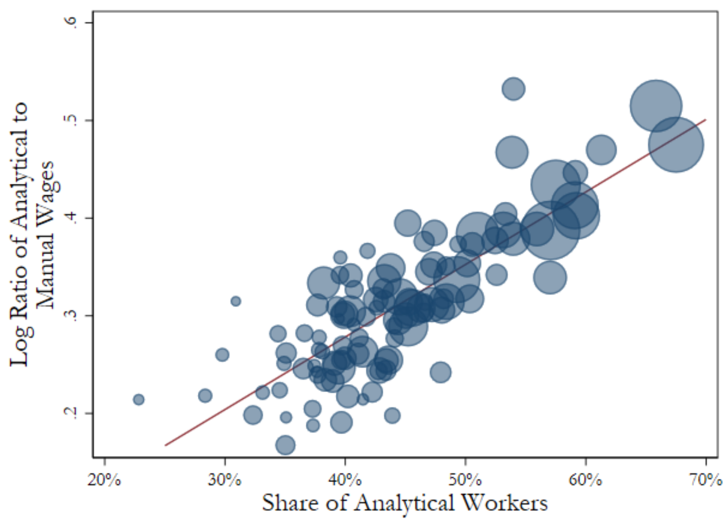
Notes: Heterogeneity in the Urban Wage Premium Across Skill and Job Task Content. Panel (a): Log annual wages by education level plotted against log population for 20 West German commuting zone quantiles. Panel (b): Log annual wages in analytical and manual occupations plotted against log population for 20 commuting zones quantiles. Bin averages are weighted by commuting-zone employment. Population corresponds to the plotted year, while cities are assigned to bins based on their 1985 population. Occupations are classified as analytical if their analytical task intensity in 1986 exceeds 0.5. The figure is constructed using population data from the German Statistical Office (DeStatis) and the Sample of Integrated Employment Biographies. For details on the sample construction, see the text. To ensure the presence of high-skill workers in 1985, the sample used for this figure imposes no restriction on workers' year of birth.

3 Theoretical Framework

In this section, we develop a model of spatial sorting that predicts that more skilled workers sort into larger cities and earn higher urban wage premia. There is a continuum of heterogeneous workers who choose their occupation and location. Workers are heterogenous in their innate ability α. We denote the density of α by $f(\alpha)$ and the corresponding support by $\mathbb{A} \equiv [\underline{\alpha}, \overline{\alpha}]$. Workers choose from a continuum of available occupations, each defined as a unique bundle of tasks, and a discrete set of locations indexed by $c \in \mathbb{C} = 1, ..., C$.

Choosing an occupation requires workers to learn the associated tasks, i.e. to accumulate human capital or skill. We assume that tasks' skill intensity increases in their average productivity. That is, workers have to accumulate more human capital to master tasks that have a higher average productivity. To produce occupation-specific output, workers solve the tasks associated with their occupation. A worker's productivity in her occupation also depends on cities' productivity advantages. Thus, locations productivity advantages and an occupation's task bundle determine a worker's productivity.

Figure 3: Relative Wages and Supply of Analytical Workers



Notes: Log ratio of analytical to manual wages in West German local labor markets versus the local share of analytical workers. Marker size proportional to local labor market population. The sample includes all male German employees born since 1960 between the ages 16 and 60 included in the SIAB working full time without having ever worked in East Germany. Population from DeStatis. Cities correspond to local labor markets as defined by Kosfeld and Werner (2012).

Two assumptions are key to our model. We assume that overall more productive locations also have a comparative advantage in the production of skill-intensive tasks and that more capable workers have a comparative advantage in learning to solve more skill-intensive tasks. We show that in our setup more productive locations end up bigger and, since they have a comparative advantage in the production of skill-intensive tasks, pay larger skill premia. It is the interaction of these comparative advantages that gives rise to spatial sorting and the skill-bias in the urban wage premium.

3.1 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 $\sigma \in \Sigma \equiv [\underline{\sigma}, \overline{\sigma}]$. Assuming a CES production function, the output of the final good is given by

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

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.

Producing occupational output requires only labor that is supplied by a mass of L heterogeneous individuals characterized by their ability α. Thus, profits of final goods producers equal output minus their wage bill

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

We define occupations as an indivisible bundle of tasks that must be completed simultaneously to produce output. In this framework, tasks can be thought of as 'a unit of work activity that produces the output' (Acemoglu and Autor, 2011). We assume that there exists a continuum of tasks, indexed by $\tau \in T \equiv [\underline{\tau}, \overline{\tau}]$. 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 occupation σ depends on the intensity at which it uses each task, on the ability α of the worker to complete the tasks, and on the location c in which it is being produced

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

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 $\int_{\tau \in \mathcal{T}} H(\tau, \sigma) \mathrm{d}\tau = 1$ for all σ.

The relative efficiency of task τ's productivity $Z(\tau, c)$, depends only on the location of production. This parameter is the only source of a given location's productive advantage and is taken as given by the worker. There are many reasons why locations might provide productive advantages that are task-specific and we do not make attempts to distinguish between them.[19] These task efficiencies

[19]Several mechanisms can give rise to this effect. Marshall (1890) identifies three key sources of agglomeration economies: labor market pooling, input sharing, and knowledge spillovers. Our 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

are allowed to differ across locations due to occupation level indivisibilities.

We 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, we also make assumptions on the functional form of occupation-specific task intensities. We 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

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

where $A(c, \sigma) = \int_{\tau \in T} Z(\tau, c) H(\tau, \sigma) \mathrm{d}\tau$ 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 mechanism on task demands is important for several reasons. First, 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. Second, it provides a means to study the evolution of occupation-specific urban wage premia following technological shocks that might alter task demands. This could be particularly relevant in explaining how location-specific returns to different occupations have evolved following technological shifts in job task requirements over past decades as observed by Autor et al. (2003).

that are task-specific.

Finally, 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 the paper, we therefore interchangeably use both the task definition of agglomeration economies and its occupation-specific equivalent, as is convenient for exposition.

Finally, we define wages. Each worker inelastically supplies one unit of labor, so her income is her physical productivity times the price of the occupational output produced,

$$w(c,\sigma;\alpha) = p(\sigma)q(c,\sigma;\alpha). \tag{5}$$

3.2 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, consider this process as a sequence of decisions, 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 a given worker's comparative advantage across occupations does not depend on their choice of location.

In the first instance, individuals choose an occupation to maximize 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

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

where $b(\tau,\alpha) > 0$ is the cost for learning a task τ by an individual of ability α. We 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 that are indexed with higher-τ, that is, in tasks that benefit more from cities' productive advantages. In Appendix A.1, we 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 one can interpret the cost $B(\sigma; \alpha)$ as an investment into education or, as is common in economics, skill. As such, this specification allows for an important distinction between skill, a property that 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 irrelevant whether we would have 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 the 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 maximize their utility, given their occupation. We assume that worker of ability α have Cobb-Douglas preferences over the consumption of the final numeraire $C(c, \sigma; \alpha)$ and housing $N(c, \sigma; \alpha)$

$$U(c, \sigma; \alpha) = C(c, \sigma; \alpha)^{1-\beta} N(c, \sigma; \alpha)^{\beta} \eta(c, \alpha) \tag{7}$$

where c denotes a worker's location, α her ability, and σ her occupation. Finally, we suppose that workers have idiosyncratic preferences across locations $\eta(c, \alpha)$ that are i.i.d. draws from a Gumbel distribution with shape parameter λ. The idiosyncratic preference shock captures (to the econometrician) unobservables that affect workers' location choices. It allows for diverging location choices of workers facing the same location-specific consumption bundles thereby acting as a dispersion force. The assumption that the preference shock follows a Gumbel distribution allows us to write location choices in logit form. Workers maximize utility is subject to the budget constraint $C(c, \sigma; \alpha) + r(c) N(c, \sigma; \alpha) \leq w(c, \sigma; \alpha)$, where $r(c)$ is the price of local housing in location c.

Given this setup, the optimization framework can be specified as follows. First, individuals choose an occupation to maximize their expected utility

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

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

$$\max_{c} U(c; \sigma(\alpha), \alpha). \tag{9}$$

3.3 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 the assumption on preferences in (7), is equal to

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

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

3.4 Definition of a Competitive Equilibrium

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

To qualify the equilibrium, we 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 locations is such that

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

and

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

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

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

where $I \equiv L\sum_c \int_\sigma \int_\alpha p(\sigma)q(\sigma,c,\alpha)\pi(c,\alpha)g(\sigma,\alpha)f(\alpha)\mathrm{d}\alpha\mathrm{d}\sigma$ is the total income of all workers in the economy. By free entry, producer 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.

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

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

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

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

3.5 Existence of a Competitive Equilibrium

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

LEMMA 1 (Occupational assignments.) *In a competitive equilibrium, there exists a continuous and strictly increasing matching function* $M:\mathbb{A}\to\Sigma$, *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.1.2. 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 that 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 the 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 we denote it simply as σ.

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

$$v(c, \sigma; \alpha) = \frac{p(\sigma)\alpha A(\sigma, c)\eta(\alpha, c)}{\beta r(c)}.$$

We exploit the fact that locational choice does not depend on the price of occupation nor on the ability of the worker (other than through the preference shock $\eta(\alpha, c)$) so that the maximization problem can be rewritten

$$\arg\max_c \ln v(c, \sigma; \alpha) = \arg\max_c \ln V(\sigma, c) + \eta(\alpha, c). \tag{17}$$

where we 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 maximization gives the share of workers of occupation σ that decide to live in city c,

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

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.[20] We 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:

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

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 characterize all endogenous variables as per (13)-(16).

In what follows, we consider the existence and uniqueness of this equilibrium. We show that a sufficient condition for existence and uniqueness is that the elasticity of substitution satisfies $\epsilon \geq 1$. In contrast, when $\epsilon < 1$, that is, when occupations are strong complements, there is no guarantee that a unique equilibrium exists.

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

[20] Note 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 $\bar{w}(c)L(c)$.

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:

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

We 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) $\dfrac{\partial F(r', r(i))}{\partial r(i)} \dfrac{\partial F(r', r(i))}{\partial r(j)} < 0$ 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 in Allen et al. (2015).[21]

To prove uniqueness, we 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 $\epsilon \geq 1$.[22] Then, uniqueness follows from Theorem 2 of Allen et al. (2015). This completes the proof of Lemma 3.

[21] 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).

[22] That $F(r', r)$ 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 $\epsilon \geq 1$.

3.6 Properties of a Competitive Equilibrium

We proceed to consider some relevant properties of this equilibrium.[23] First, we 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:

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

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 captures congestion costs. The lower dispersion in preferences (higher λ) and the lower the share of spending on housing, the more likely it is that agglomeration forces out-power congestion forces and guarantees that location size is increasing in location TFP. If congestion costs are overbearing however, it is possible to observe locations that are at the top of productivity distribution, but smaller in size, as rents crowd out low-income workers. One such example would be, for

[23] Note 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 $0 < \epsilon < \infty$.

instance, Silicon Valley, which has some of the highest average wages, but low population density due to high real estate prices.

We 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 the city's population as well. In the subsequent chapter, we provide empirical evidence consistent with this assumption.

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

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

where $\tilde{\kappa}(\sigma)$ is again a collection of variables that are independent of c.[24] 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:

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

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:

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

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

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

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 that benefit the most from location-specific production advantages 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.*[25]

This complementarity in ability and city size also drives the sorting of workers across locations of different sizes, as the 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 $c' > c$:

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

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

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

4 Heterogeneity of the Urban Wage Premium and Spatial Sorting

Relative to existing frameworks, our model delivers two novel predictions. First, the urban wage premium is higher for occupations with more skill-intensive task content. Second, such occupations sort more strongly into larger cities. Using German administrative employee data, we test Proposition 1, which states that the urban wage premium is heterogeneous across tasks, and Proposition 2, which states that larger cities are relatively more intensive in tasks that benefit more from urban productivity advantages.

4.1 Task-specific Urban Wage Premia

4.1.1 Baseline Specification

First, we test whether there exist significant differences in urban wage gains across tasks. Motivated by equation (21), we estimate the following specification:

$$\log w_{i,t} = \gamma \log pop_{c(i,t)} + \sum_{\tau \in T} \gamma_\tau h^{\tau}_{\sigma(i,t)} \times \log pop_{c(i,t)} + x'_{i,t}\beta + \varepsilon_{i,t}, \tag{22}$$

where $w_{i,t}$ denotes wages of worker i in year t and $x_{i,t}$ are time-varying worker characteristics such as (squared) experience and tenure (at establishment). In addition, we include worker, year, and 2-digit industry fixed effects but omit them from equation (22) for brevity. Finally, $h^{\tau}_{\sigma(i,t)}$ denotes occupation $\sigma(i,t)$'s *standardized* task intensity in task group $\tau \in T$. The descriptive patterns suggest a difference in the urban wage premium between manual and analytical tasks. Thus, we interact log population with the standardized analytical task intensity of a worker's current occupation.[26]

The parameters of interest are γ, which measures the wage elasticity with respect to city population *pop*, and γ_τ, which measures how much the wage elasticity with respect to city population

[26]We do not interact the manual task intensity with local population, because, by construction, an occupation's manual task intensity equals one minus its analytical task intensity.

changes with worker i's current occupation's task content. Put differently, γ_τ measures task-specific urban productivity gains. Because a city's population varies only marginally over time, γ is primarily identified from workers who move between cities. In contrast, γ_τ is primarily identified from workers moving between cities and switching occupations. Motivated by the descriptive patterns from Section 2, we focus on how the urban wage premium varies with an occupation's analytical task intensity.

A few comments are in order. The literature usually estimates the effect of population on wages in two stages to ensure that standard errors are unbiased.[27] We deviate from the existing literature and bootstrap our standard errors using 1,000 bootstrap samples, because we interact city-level variables with individual characteristics in most of our specifications. Second, we do not control for the specific task intensities, since we standardize (and therefore demean) our analytical task intensity. Hence, γ_τ is still interpreted as marginal effect for a worker currently employed in an occupation with a mean analytical task intensity. We do not include occupation fixed effects, since γ_τ would no longer be estimated using variation of workers that switch between occupation groups.

To compare our results to those of other studies, we also estimate the wage elasticity with respect to population without allowing for heterogeneity across tasks. That is, we omit the interaction term $h^{\tau}_{\sigma(i,t)} \times \log pop_{c(i,t)}$ from specification (22). Table 2 presents the results from this exercise. Column (1) provides estimates without heterogeneity across tasks omitting worker fixed effects. Column (2) includes the estimates in a specification with worker fixed effects. We estimate that the wage elasticity is around 4% without and 1.8% with worker fixed effects. These lie within the range of the preceding literature. That the inclusion of worker fixed effects reduces the wage elasticity by roughly 50% is also in line with previous studies.[28]

In Table 3, we present the results from our baseline regression that allows for a task-bias in the urban wage premium. The point estimates suggest a strong role for tasks in determining the magnitude of the urban wage premium. Note that since we standardize our task intensity measure

[27] In a nutshell, jointly estimating group, e.g. city, and worker level variables, e.g. job task content, can lead to biased standard errors (Moulton, 1990). Since the parameter of interest γ is identified from workers moving between cities, clustering standard errors is not a suitable solution. Instead, one estimates city fixed effects while controlling for worker level variables in a first stage and then regresses the city fixed effects on city population (Combes et al., 2008).

[28] In table C.8, we present the corresponding estimates from (Combes et al., 2008)'s two-stage estimation.

Table 2: Conditional Urban Wage Premium without Heterogeneity

Dependent variable:	*Log daily wages*	
	(1)	(2)
Log population	0.0424***	0.0180***
	(0.0009)	(0.0013)
Experience	0.0268***	0.0414***
	(0.0003)	(0.0012)
Experience^2	-0.0425***	-0.0873***
	(0.0007)	(0.0006)
Tenure	0.0140***	0.0006***
	(0.0002)	(0.0001)
Tenure^2	-0.0297***	0.0009*
	(0.0007)	(0.0005)
Vocational Degree	0.1623***	
	(0.0025)	
College Degree	0.6805***	
	(0.0034)	
Constant	3.5398***	4.1240
	(0.0129)	(0.0256)
Fixed Effects:		
Worker		✓
Year	✓	✓
2-digit industry	✓	✓
No. of Observations	1,443,091	1,443,091
Adjusted Within R^2	0.2812	0.0431

Notes: Notes: The specifications include a constant as well as person, year, two-digit industry, experience and quadratic experience. Standard errors are bootstrapped using 1,000 clustered bootstrap samples. ***, **, and * indicate significance at the 0.1, 1, and 5 percent levels. We report adjusted R^2 within workers.

and we include log population as a separate, γ_τ measures the change in the wage elasticity with respect to local population as a result of a one standard deviation increase in the worker's current occupation's analytical task intensity. Thus, increasing an occupation's analytical task intensity by one standard deviation increases the wage elasticity with respect to local population by 0.18 percentage points (8% in relative terms). For the least analytical occupation, i.e. train conductors, the urban wage premium is a mere 1.7%. In contrast, for the most analytical occupation, e.g. jobs in marketing, the conditional urban wage premium increases to 2.3%.

Table 3: Estimation of Task-Specific Population Earnings Premia

Dependent variable:	*Log daily wages*	
	(1)	(2)
Log population	0.0207***	0.0064***
	(0.0010)	(0.0014)
Log population × analytical intensity	0.0016***	0.0015**
	(0.0001)	(0.0001)
Log population × vocational dummy		.0107***
		(0.0003)
Log population × college dummy		0.0241***
		(0.0005)
N	1,443,091	1,443,091
Ajusted Within R^2	0.0444	0.0529

Notes: Notes: The specifications include a constant as well as person, year, two-digit industry, experience and quadratic experience. Standard errors are bootstrapped using 1,000 clustered bootstrap samples. ***, **, and * indicate significance at the 0.1, 1, and 5 percent levels. We report adjusted R^2 within workers.

4.1.2 Task or Skill Bias?

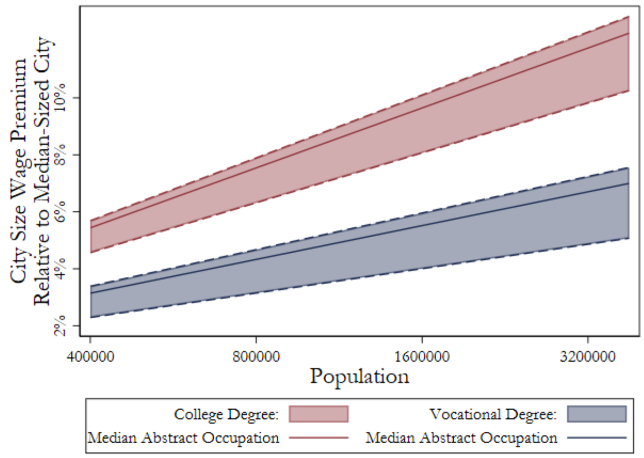
In our sample, workers in analytical occupations are more likely to have a college degree (29% vs 4.6%). Thus, a concern with specification (22) is that our task-based approach captures the skill-bias of the urban wage premium. If, conditionally on our controls and fixed effects, educational attainment and analytical task intensity were positively correlated, then the estimates of γ_τ would be upward biased. To control for this confounding effect, we include interactions between educational attainment dummies and log population, i.e.,

$$\begin{aligned} \log w_{i,t} = \gamma \log pop_{c(i,t)} + \sum_{\tau \in T} \gamma_\tau h^{\tau}_{\sigma(i,t)} \times \log pop_{c(i,t)} \\ + \sum_{s \in S} \gamma_s \mathbb{1}_s \times \log pop_{c(i,t)} + x'_{i,t}\beta + \varepsilon_{i,t}, \end{aligned} \tag{23}$$

where variables are defined as before and $\mathbb{1}_s$ is an indicator function equal to 1 if worker i has education level $s \in \{\text{Vocational Degree}, \text{ College Degree}\}$. The base level is *At Most School* which in most cases corresponds to 10 years of schooling.[29] We omit education dummies because of the

[29]Note that the SIAB does not include the number of years in school, but only specifies the highest degree a workers has attained.

Figure 4: The City Size Wage Premium By Task and Education



Notes: Predicted city size wage premium by education and analytical task intensity for workers living in cities larger than median-size relative to a median-sized city (400,000 residents). For a given population, the area shaded in red (blue) gives the range of the relative wage premium for those with a college (vocational) degree depending on their occupation's task intensity. Prediction is based on the estimation of specification 23. Coefficient values correspond to the estimates presented in Table 3.

inclusion of worker fixed effects. In this specification, the parameter γ_s measures the variation in the urban wage premium across tasks not captured by variation in educational attainment.

Column (2) of Table 3 presents the estimates. The conditional wage elasticity with respect to local population for workers without college or vocational degrees in an occupation with average analytical task intensity is 0.6% and increases to 3% (1.6%) for workers with college (vocational) degrees. Crucially though, the urban wage premium is still stronger for more analytical occupations. In fact, the coefficient on the interaction of log population and analytical task intensity barely changes, dropping from 0.0016 to 0.0015. We take this as evidence that the task-bias exists even within skill groups. Figure 4 emphasizes the former point.

4.2 Spatial Sorting by Task Content

In our model, the heterogeneity of the urban wage premium across tasks generates spatial sorting of occupations. Proposition II 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

$$\ln\left(\frac{\pi(\sigma',c)}{\pi(\sigma,c)}\right) = \frac{1}{\lambda}\int_{\tau\in T} Z(\tau,c)\left[H(\tau,\sigma') - H(\tau,\sigma)\right]\mathrm{d}\tau \\ + \ln\frac{\sum_k (e^{A(\sigma,k)}r(k)^{-\beta})^{1/\lambda}}{\sum_k (e^{A(\sigma',k)}r(k)^{-\beta})^{1/\lambda}}. \tag{24}$$

Building on Lemma 2, Proposition 2 states that the distribution of workers is log-supermodular in worker ability and city size. As a result of higher-ability workers sorting into more analytical occupations, this implies that relatively more analytical occupations should be relatively more common in larger cities. We test the conclusion of Proposition II, i.e. sorting of occupations, by estimating (24) with a cross-section of observed differences in sorting into some location c between each occupation-pair σ and σ'

$$\ln\left(\frac{\pi(\sigma',c)}{\pi(\sigma,c)}\right) = \nu_{\sigma,\sigma'} + \sum_{\tau}\theta_\tau\left(h^\tau_\sigma - h^\tau_{\sigma'}\right)\times\log pop_c + \epsilon_{\sigma\sigma'c}, \tag{25}$$

where $\pi(\sigma,c)$ is the fraction of workers in occupation σ working in location c; $\nu_{\sigma,\sigma'}$ are occupation-pair fixed effects for occupations σ and σ' and $\epsilon_{\sigma\sigma'c}$ denotes the error term. The fixed effect $\nu_{\sigma,\sigma'}$ controls for the term

$$\ln\left(\frac{\left(\sum_k (e^{A(\sigma,k)}r(k)^{-\beta})^{1/\lambda}\right)}{\left(\sum_k (e^{A(\sigma',k)}r(k)^{-\beta})^{1/\lambda}\right)}\right), \tag{26}$$

in equation (24).

The parameter θ_τ measures the spatial sorting of workers across tasks. It captures the extent to which differences in task intensities between two occupations correlate with their relative sorting rates into larger cities. We expect the estimates of θ_τ to be correlated with the estimates of γ_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.

To facilitate the interpretation of the parameter θ_τ, we standardize the term $(h^\tau_\sigma - h^\tau_{\sigma'})$ and log population subtracting their mean and dividing each variable by its standard deviation. Table 4

Table 4: Worker Sorting By Tasks

Dependent Variable:	$\ln\left(\frac{\pi(\sigma',c)}{\pi(\sigma,c)}\right)$
$(h_{analytical,\sigma'} - h_{analytical,\sigma}) \times \log$ population	2.4924***
	(0.0791)
N	1,220,963
R^2	0.3344

Notes: Estimates for model 25 occupation-pair fixed effects. Coefficients are reported with heteroscedastic robust standard errors in parenthesis. ***, **, and * indicate significance at the 0.1, 1, and 5 percent levels. We report adjusted R^2.

presents the OLS estimates of equation (26). θ_τ equals 2.4924 and is strongly statistically significant.

4.3 Sensitivity

We now examine the robustness of our results. First, we explore whether our results are robust to our sample selection criteria. Second, we explore the sensitivity of our results to using alternative labor market definitions. Throughout, we focus on our results on the task-bias in the urban wage premium and the relative sorting of occupations by their task content into larger cities. The results from our robustness analysis for the task-bias in the urban wage premium and the sorting by tasks are presented in tables C.9 and C.10, respectively.

In our analysis, we focused on male employees in West Germany.[30] We now repeat our analysis using a sample of female workers in West Germany and male workers in East Germany.[31] When using the female sample, we do not change other sample selection criteria, e.g. birth year and sample period. In contrast, because workers in East Germany are only observed after 1993, we require that subjects in the sample are born no earlier than 1975. This ensures that we observe each worker's full employment biographies, i.e. that workers are no older than 18 years in 1993.

For female workers in West Germany, the estimates display similar patterns to those for men in West Germany, although the variation in the urban wage premium is slightly larger. Including

[30] We restrict our sample to workers in West Germany for two reasons. First, there still exist substantial structural economic differences between East and West Germany. Secondly, the data for West Germany includes information on employment spells (relevant for experience) back to 1975 and on wages since 1985. Thus, restricting our sample to West Germany also increases the age range of our workers.

[31] Five (out of 72) districts in East Germany are part of West German local labor markets (i.e., Vogtlandkreis, Salzwedel, Harz, Eichsfeld, and Sonneberg). Because workers in these districts earn West German wages, we drop them.

interactions between log population and education dummies reduces the base elasticity but does not substantially affect the coefficient quantifying the task-bias. The results suggest that the task-bias in the urban wage premium takes a similar form for both men and women, but that the overall urban wage premium is somewhat larger for women. This could be explained by a variety of factors such as differing employment choice between rural and urban women. Better access to childcare in cities may allow urban women to work managerial positions more easily.

For male workers in East Germany, we observe no urban wage premium for workers with an average analytical task intensity (i.e. γ is not statistically significantly different from zero), but some effect for highly analytical occupations. That is, the interaction term of the standardized analytical task intensity is statistically significantly different from zero, although lower than for men in West Germany. The urban wage premium may be significantly smaller in East Germany because of fewer large firms in large East German cities. Recent studies suggest that the sorting of large firms into large cities may partially explain the urban wage premium (e.g. Gaubert, 2018). However, following the second World War large private firms moved to West German locations. Following Germany's reunification most large public companies went bankrupt. As a result, there are virtually no large production sites or firms in East German agglomeration. The smaller coefficient on the interaction of log population and analytical task intensity could be explained by differences in the East and West German samples. The East German sample is younger due to more restrictive birth year restrictions. Thus, if differences in the urban wage premium across occupational task content arise only over time, e.g., because workers in analytical occupations benefit more from learning externalities in larger cities, then this could explain the disparities in the results. Nevertheless, since the qualitative results are broadly similar for women in West Germany and men in East Germany, we believe that our empirical results are robust to alternative sample selection criteria.

To address commuting between districts, we analyze spatial sorting across labor markets and not individual cities as defined by (Kosfeld and Werner, 2012). This definition minimizes commuting between our units of observation but results in very large labor market definitions. We now explore whether our results are robust to using a more granular labor market definition by the Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR). We obey the sample selection criteria from our main analysis, in particular, we restrict the sample to West Germany. Using the alternative labor market definition from the BBSR does not change our conclusions.

5 Conclusion

Since the 1980s, high-skill workers increasingly sort into larger cities. The existing literature explains this development by noting that the urban wage premium has become (relatively) larger for high-skill workers. We propose a theoretical framework with a novel microfoundation for the skill-bias in the urban wage premium and, by extension, the spatial sorting by skill level. In our model, larger cities end up having a comparative advantage for skill-intensive occupations. The strength of this comparative advantage is driven by occupational task content and heterogeneity in cities' comparative advantage in skill-intensive tasks. A key insight of our model is that changes in occupational task content, perhaps caused by the introduction of new technologies, can explain changes in spatial sorting over time.

We find strong support for the properties of our model's equilibrium in the data. Using rich administrative data from Germany, we show that the urban wage premium differs significantly across occupations and, in particular, is largest for jobs intensive in analytical tasks. 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 in the quest to disentangle the 'black' box that is the urban productivity gains.

Our results highlight that differences in the urban wage premium across occupations can potentially explain the skill-bias and spatial sorting, but we fall short of explaining why larger cities may have a comparative advantage in skill-intensive occupations. Thus, future research should explore potential mechanisms explaining why the urban wage premium varies across occupations.

References

Acemoglu, Daron and David Autor, "Skills, Tasks and Technologies: Implications for Employment and Earnings," in "Handbook of Labor Economics," Elsevier, 2011, pp. 1043–1171.

Adda, Jerome and Christian Dustmann, "Sources of Wage Growth," *Journal of Political Economy*, February 2023, *131* (2), 456–503.

Allen, Treb, Costas Arkolakis, and Xiangliang Li, "On the existence and uniqueness of trade equilibria," *Manuscript, Yale Univ*, 2015.

Autor, D. H., F. Levy, and R. J. Murnane, "The Skill Content of Recent Technological Change: An Empirical Exploration," *The Quarterly Journal of Economics*, November 2003, *118* (4), 1279–1333.

Autor, David, "The Faltering Escalator of Urban Opportunity," in Melissa S Kearney and Amy Ganz, eds., *Securing Our Economic Future*, Aspen Institute, December 2020, p. 108–36.

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

Baum-Snow, N. and R. Pavan, "Understanding the City Size Wage Gap," *The Review of Economic Studies*, August 2011, *79* (1), 88–127.

Baum-Snow, Nathaniel and Ronni Pavan, "Inequality and City Size," *The Review of Economics and Statistics*, December 2013, *95* (5), 1535–1548.

_ , Matthew Freedman, and Ronni Pavan, "Why Has Urban Inequality Increased?," *American Economic Journal: Applied Economics*, October 2018, *10* (4), 1–42.

Becker, Gary, *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education, First Edition*, National Bureau of Economic Research, Inc, 1964.

Behrens, Kristian and Frédéric Robert-Nicoud, "Chapter 4 - Agglomeration Theory with Heterogeneous Agents," in Gilles Duranton, J. Vernon Henderson, and William C. Strange, eds., *Handbook of Regional and Urban Economics*, Vol. 5 of *Handbook of Regional and Urban Economics*, Elsevier, 2015, pp. 171–245.

_ , Gilles Duranton, and Frédéric Robert-Nicoud, "Productive Cities: Sorting, Selection, and Agglomeration," *Journal of Political Economy*, June 2014, *122* (3), 507–553.

Berry, Christopher R. and Edward L. Glaeser, "The divergence of human capital levels across cities," *Papers in Regional Science*, August 2005, *84* (3), 407–444.

Cerina, Fabio, Elisa Dienesch, Alessio Moro, and Michelle Rendall, "Spatial Polarisation," *The Economic Journal*, July 2022, *133* (649), 30–69.

Ciccone, Antonio and Robert Hall, "Productivity and the Density of Economic Activity," *American Economic Review*, 1996, *86* (1), 54–70.

Combes, Pierre-Philippe and Laurent Gobillon, "The Empirics of Agglomeration Economies," in "Handbook of Regional and Urban Economics," Elsevier, 2015, pp. 247–348.

__ , **Gilles Duranton, and Laurent Gobillon**, "Spatial wage disparities: Sorting matters!," *Journal of Urban Economics*, March 2008, *63* (2), 723–742.

Costinot, Arnaud and Jonathan Vogel, "Matching and Inequality in the World Economy," *Journal of Political Economy*, August 2010, *118* (4), 747–786.

__ **and** __ , "Beyond Ricardo: Assignment Models in International Trade," *Annual Review of Economics*, August 2015, *7* (1), 31–62.

Dauth, Wolfgang and Johann Eppelsheimer, "Preparing the sample of integrated labour market biographies (SIAB) for scientific analysis: a guide," *Journal for Labour Market Research*, August 2021, *54* (1).

Davis, Donald R. and Jonathan I. Dingel, "A Spatial Knowledge Economy," *American Economic Review*, January 2019, *109* (1), 153–170.

__ **and** __ , "The comparative advantage of cities," *Journal of International Economics*, March 2020, *123*, 103291.

de la Roca, Jorge and Diego Puga, "Learning by Working in Big Cities," *The Review of Economic Studies*, 07 2017, *84* (1), 106–142.

Dengler, Katharina, Britta Matthes, and Wiebke Paulus, "Berufliche Tasks auf dem deutschen Arbeitsmarkt: eine alternative Messung auf Basis einer Expertendatenbank.," *FDZ-Methodenreport, 12/2014 (de)*, 2014.

Diamond, Rebecca, "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000," *American Economic Review*, March 2016, *106* (3), 479–524.

__ **and Cecile Gaubert**, "Spatial Sorting and Inequality," *Annual Review of Economics*, 2022, *14* (1), 795–819.

__ **and Juan Carlos Suárez Serrato**, "Spatial Sorting and Inequality," Working Paper 33609, National Bureau of Economic Research March 2025.

Eckert, Fabian, Mads Hejlesen, and Conor Walsh, "The return to big-city experience: Evidence from refugees in Denmark," *Journal of Urban Economics*, 2022, p. 103454.

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

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

Ganzer, Andreas; Schmucker Alexandra; Vom Berge Philipp Frodermann Corinna;, "Stichprobe der Integrierten Arbeitsmarktbiografien Regionalfile (SIAB-R) 1975-2019.," *FDZ-Datenreport, 05/2021 (en)*, 2021.

Gaubert, Cecile, "Firm Sorting and Agglomeration," *American Economic Review*, November 2018, *108* (11), 3117–53.

Gaubert, Cécile and Frédéric Robert-Nicoud, “Sorting to Expensive Cities,” Working Paper 33652, National Bureau of Economic Research April 2025.

Giannone, Elisa, “Skill-Biased Technical Change and Regional Convergence,” Working Paper, Centre de Recerca en Economia Internacional (CREI) jul 2022. Revised July 2022; original March 2021.

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

Gould, Eric D., “Cities, Workers, and Wages: A Structural Analysis of the Urban Wage Premium,” *The Review of Economic Studies*, April 2007, *74* (2), 477–506.

Grujovic, Anja, “Tasks, cities and urban wage premia,” *CEPREMAP Working Papers 1807*, November 2018.

Kosfeld, Reinhold and Alexander Werner, “German Labour Markets—New Delineation after the Reforms of German District Boundaries 2007–2011,” *Raumforschung und Raumordnung / Spatial Research and Planning*, Feb. 2012, *70* (1), 49–64.

Koster, Hans R.A. and Ceren Ozgen, “Cities and tasks,” *Journal of Urban Economics*, November 2021, *126*, 103386.

Marshall, Alfred, Macmillan, London, 1890.

Michaels, Guy, Ferdinand Rauch, and Stephen J Redding, “Task Specialization in U.S. Cities from 1880 to 2000,” *Journal of the European Economic Association*, March 2018, *17* (3), 754–798.

Moretti, Enrico, “Real Wage Inequality,” *American Economic Journal: Applied Economics*, January 2013, *5* (1), 65–103.

Moulton, Brent R., “An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units,” *The Review of Economics and Statistics*, 1990, *72* (2), 334–338.

Parkhomenko, Andrii, “Homeownership, Polarization, and Inequality,” *The Review of Economic Studies*, 09 2025, p. rdaf068.

Perl, Maximilian, *Agglomerations, Tasks and Wage Growth*, DE: RWI, 2023.

Rosenthal, Stuart S. and William C. Strange, “Chapter 49 Evidence on the nature and sources of agglomeration economies,” in “Handbook of Regional and Urban Economics,” Elsevier, 2004, pp. 2119–2171.

A Theory Appendix

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

By definition:

$$B(\sigma; \alpha) \equiv \int_{\tau \in T} b(\tau, \alpha) H(\tau, \sigma) \mathrm{d}\tau \tag{27}$$

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 $\int_{\tau \in \tau} H(\tau, \sigma) \mathrm{d}\tau = 1$ for all σ.

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

$$\begin{aligned}
B(\sigma, \alpha) + B(\sigma', \alpha') &< B(\sigma', \alpha) + B(\sigma, \alpha') \\
\int_{\tau \in T} b(\tau, \alpha) H(\tau, \sigma) \mathrm{d}\tau + \int_{\tau \in T} b(\tau, \alpha') H(\tau, \sigma') \mathrm{d}\tau &< \int_{\tau \in T} b(\tau, \alpha) H(\tau, \sigma') \mathrm{d}\tau + \int_{\tau \in T} b(\tau, \alpha') H(\tau, \sigma) \mathrm{d}\tau \\
\int_{\tau \in T} [b(\tau, \alpha) - b(\tau, \alpha')][H(\tau, \sigma) - H(\tau, \sigma')] \mathrm{d}\tau &< 0
\end{aligned} \tag{28}$$

By assumption of monotonicity on $H(\tau, \sigma)$, its second order cross derivative, and by assumption $\int_{\tau \in \tau} H(\tau, \sigma) \mathrm{d}\tau = 1$ 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 (28) can be rewritten:

$$\int_{\tau_{\min}}^{\bar{\tau}} \underbrace{[b(\tau, \alpha) - b(\tau, \alpha')]}_{>0 \, \forall \, \tau} \underbrace{[H(\tau, \sigma) - H(\tau, \sigma')]}_{\geq 0 \, \forall \, \tau} \mathrm{d}\tau < \int_{\bar{\tau}}^{\tau_{\max}} \underbrace{[b(\tau, \alpha) - b(\tau, \alpha')]}_{>0 \, \forall \, \tau} \underbrace{[H(\tau, \sigma') - H(\tau, \sigma)]}_{\geq 0 \, \forall \, \tau} \mathrm{d}\tau \tag{29}$$

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

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

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

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

Therefore, the left-hand side of (29) must strictly be lower than the right hand side, which implies that the inequality in (28) 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 our 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 our 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 we will use in this proof can be shown to be the following, given that the the error terms are distributed Type I extreme value:

$$\begin{aligned}
\ln V(\sigma;\alpha) &\equiv \mathrm{E}[\max_c V(\sigma,c;\alpha)] \\
&= \lambda \ln\left[\sum_c V(\sigma,c;\alpha)^{1/\lambda}\right] \qquad \text{(as shown by Rosen and Small, 1981)} \\
&= \lambda \ln\left[\sum_c \left(\frac{p(\sigma)e^{\alpha+A(\sigma,c)}}{r(c)^\beta}\right)^{1/\lambda}\right] \\
&= \ln(p(\sigma)e^\alpha) + \ln\left[\sum_c \left(\frac{e^{A(\sigma,c)}}{r(c)^\beta}\right)^{1/\lambda}\right] \\
&= \ln\left[p(\sigma)e^\alpha \tilde{A}(\sigma)\right] \qquad \text{where} \qquad \tilde{A}(\sigma) \equiv \sum_c \left(\frac{e^{A(\sigma,c)}}{r(c)^\beta}\right)^{1/\lambda} \qquad (30)
\end{aligned}$$

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

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

Conditions (15) and (16), together with $f(\alpha) > 0$ for all α, imply that that $\Sigma(\alpha) \neq \varnothing$ for all $\alpha \in \mathbb{A}$. To show that $\mathbb{A}(\sigma) \neq \varnothing$ for all σ, we proceed by contradiction. Suppose that there exists σ' such that $\mathbb{A}(\sigma') = \varnothing$. Since $\Sigma(\alpha) \neq \varnothing$ for all $\alpha \in \mathbb{A}$, we know that there exists σ such that $\mathbb{A}(\sigma) \neq \varnothing$. 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. $\mathbb{A}(.)$ 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 $\alpha' \geq \alpha$.

We 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, we 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 $\Sigma(\alpha_2) \neq \varnothing$ 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, we 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 $H : \Sigma \to \mathbb{A}$ 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 $\Sigma(\alpha) \neq \varnothing$ 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 $H : \Sigma \to \mathbb{A}$ 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 Sample of Integrated Employment Biographies

Our principle data source is the sample of integrated employment biographies (SIAB) covering the years 1975 through 2019. For a detailed description of the version used in this paper, see Frodermann (2021). Data access is provided through the IAB Data Research Center (IAB-FDZ). We applied for the following additional variable: district of employing establishment ("ao_kreis").

For our main analysis, we restrict the sample to employment spells between 1985 and 2010 of males employed full-time in private-sector industries who are German nationals residing in West Germany, aged 16 to 65, and born in or after 1969.

We exclude part-time and marginally employed workers, because the SIAB does not report the hours worked, preventing us from computing full-time equivalent wages.[32]

We also exclude workers in the primary sector, whose employment is heavily influenced by natural resource availability. We ensure that we observe workers' entire labor market biographies by dropping foreign nationals and individuals born before 1969 (i.e. those no older than 16 in 1975). We limit the analysis to the period 1985–2010, as our task intensity measures map consistently into the 1985 German occupational classification; from 2010 onward, these values are imputed based on a different classification system.

We further restrict the sample to West Germany, excluding workers with any employment spell in East Germany or West Berlin, to maximize the study period and because of fundamental differences between the East and West German economy in the 1990s.[33]

[32] Marginal employment (*geringfügig beschäftigt*) refers to jobs with monthly earnings below a statutory threshold or no more than 70 working days per calendar year. As of 2019, this threshold was €450.

[33] Much of the Berlin local labor market includes areas formerly part of East Germany. Moreover, prior to reunification, Berlin was heavily subsidized by the federal government, distorting labor market incentives.

C Further Figures and Tables

Table C.5: Classification of BIBB task items into aggregate task groups

Purchasing, procuring, selling	Analytical
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	
Measuring, testing, quality control	
Monitoring, control of machines, plants, technical processes	
Repairing, refurbishing	Manual
Entertaining, accommodating, preparing food	
Nursing, caring, healing	
Protecting, guarding, patrolling, directing traffic	
Cleaning, removing waste, recycling	
Transporting, storing, shipping	

Notes: Table shows classification of tasks from BIBB QCS 1986 wave into two aggregate task groups following Adda and Dustmann (2023).

Table C.6: Analytical Task Intensity By Skill

	Mean Analytical Task Intensity
Low Skill Job	0.13
Medium Skill Job	0.37
High Skill Job	0.86
Very High Skill Job	0.91
No College	0.40
College	0.88

Notes: Notes: Mean analytical task intensity by jobs' skill requirements and worker education for individuals in SIAB. Observations are worker-year pairs. Data is for male employees that are (i) working full-time, (ii) aged 18 to 65, (iii) German nationals, (iv) never worked in East Germany and are contained in the Sample of Integrated Employment Biographies (SIAB).

Table C.7: Top and Bottom Five Occupations By Task Intensity

	Analytical Non-Routine	*Manual Non-Routine*	*Manual Routine*	*Interactive Non-Routine*	*Cognitive Routine*
Top 5					
1	Humanities	Train Driver & Traffic Control	Ceramic Manufacturing	Presenters & Entertainers	Tax Consulting
2	Mathematics & Statistics	Building Construction	Plastic & Rubber Manufacturing	Marketing	Mechatronics & Automation
3	Language- & Literature Studies	Underground Construction	Glass manufacturing	Book, Art, Antique & Music Stores	Electrical Engineering
4	Veterinarians & Related	Control & Maintenance in Transport	Printing & Bookbinding	Service staff in Passenger Transport	Technical Drawing & Related
5	Geology, Geography & Meteorology	Jobs in Body Care	Metal Production	Education & Social Work	Auditing & Accounting
Bottom 5					
136	Metal Production	Pharmacy	Biologist	Ceramic production/processing	Driving & Sports Instructors
137	Floor Installation	Tax Consulting	Tax Consulting	Metal Construction/welding	Teaching & Research at Universities
138	Operators of Construction Machinery	Accounting & Auditing	Computer Science	Stage and Costume Design	Veterinarians & Related
139	Ceramic Production/Processing	Finance & Insurance	Presenters & Entertainers	Beverage Production	Train Drivers & Traffic Control
140	Train Drivers & Traffic Control	Marketing	Marketing	Paper & Packaging Technology	Actors & Dancers

Notes: The table lists 5 occupations with the lowest and the highest task intensity for each of the five task groups constructed by Dengler et al. (2014). See section 2 for details.

Figure C.5: The Skill-bias of the Urban Wage Premium

Notes: Difference between log annual wages for college graduates and workers without a college degree plotted against log population for 20 West German commuting zone quantiles. Commuting zones were grouped into 20 quantiles based on their population to satisfy data protection requirements. The figure is constructed using population data from the German statistical office (DeStatis) and data on wages and education is for male employees that are (i) working full-time, (ii) aged 18 to 65, (iii) German nationals, (iv) never worked in East Germany and are contained in the Sample of Integrated Employment Biographies (SIAB). Note that there is no birth year restriction on workers contained in the data to ensure that there are high-skill workers in 1985.

Table C.8: Conditional Urban Wage Premium without Heterogeneity: Two-stage Estimation

Dependent variable:	*Log daily wages*	*City fixed effects*	*Log daily wages*	*City fixed effects*
	(1)	(2)	(3)	(4)
Log population		0.0426***		0.0112***
		(0.0073)		(0.0038)
Experience	0.0388***		0.0388***	
	(0.0003)		(0.0003)	
Experience^2	-0.0740***		-0.0740***	
	(0.0008)		(0.0008)	
Tenure	0.0110***		0.0110***	
	(0.0002)		(0.0002)	
Tenure^2	-0.0257***		-0.0257***	
	(0.0008)		(0.0009)	
Fixed Effects				
Worker			✓	
Year	✓		✓	
2-digit Industry	✓		✓	
City	✓		✓	
Education	✓			
No. of Observations	1,443,091	107	1,443,091	107
R^2	0.3266		0.0440	

Notes: Two-stage estimation of the conditional urban wage premium following Combes et al. (2008). Column (1) presents the results for the first stage with log daily wages as dependent and city fixed effects as independent variables. The specification includes a constant as well as person, year, two-digit industry, experience and quadratic experience. Column (2) and (3) presents the estimates of regressing the city fixed effects on city population in 2010 and a constant. Column (2) uses OLS. In column (3), we instrument for contemporary population with commuting zone population from the 1860s. Standard errors are heteroscedastic robust and, in columns (1) and (3), are clustered at the worker level. ***, **, and * indicate significance at the 0.1, 1, and 5 percent levels. We report adjusted R^2 within workers.

Table C.9: Sensitivy Checks: Task-specific urban earnings premia

Sensitivity Checks	*Dependent variable:*	*Log daily wages*	
	Independent variable:	(1)	(2)
Females in West Germany	Log population	0.0256***	0.0102***
		(0.0018)	(0.0018)
	Log population × analytical intensity	0.0022**	0.0020***
		(0.0001)	(0.0001)
Males in East Germany	Log population	0.0107***	0.0016
		(0.0028)	(0.0029)
	Log population × analytical intensity	0.0007***	0.0007***
		(0.0002)	(0.0002)
BBSR Labor Market Definition	Log population	0.0177***	0.0057***
		(0.0012)	(0.0013)
	Log population × analytical intensity	0.0015***	0.0014***
		(0.0001)	(0.0001)
Controls for:			
	Log population × College Dummy		✓
	Log population × Vocational Dummy		✓

Notes: Results of our sensitivity analysis for our main model testing for a task-bias in the urban wage premium. The specification includes a constant as well as person, year, two-digit industry, experience and quadratic experience. Standard errors are bootstrapped using 1000 clustered bootstrap samples. ***, **, and * indicate significance at the 0.1, 1, and 5 percent levels. We report adjusted R^2 within workers.

Table C.10: Sensitivity Checks: Worker Sorting By Tasks

	Dependent Variable:	$\ln\left(\frac{\pi(\sigma',c)}{\pi(\sigma,c)}\right)$
Females in West Germany	$(h_{analytical,\sigma'} - h_{analytical,\sigma})\times$ log population	7.5423*** (0.1825)
Males in East Germany	$(h_{analytical,\sigma'} - h_{analytical,\sigma})\times$ log population	1.353*** (0.1017)
BBSR Labor Market Definition	$(h_{analytical,\sigma'} - h_{analytical,\sigma})\times$ log population	1.8877*** (0.0671)

Notes: Estimates for model 25 occupation-pair fixed effects. Coefficients are reported with heteroscedastic robust standard errors in parenthesis. ***, **, and * indicate significance at the 0.1, 1, and 5 percent levels. We report adjusted R^2.