

# Task Specialization in U.S. Cities from 1880-2000\*

Guy Michaels

*London School of Economics and CEP*<sup>†</sup>

Ferdinand Rauch

*University of Oxford and CEP*<sup>‡</sup>

Stephen J. Redding

*Princeton University and CEP*<sup>§</sup>

December 15, 2013

## Abstract

We introduce a new methodology for measuring the production tasks undertaken within occupations. We use this methodology to provide the first evidence on the evolution of production tasks in urban and rural areas in the United States from 1880-2000. We find that tasks involving thought, communication and intersocial activity (“interactiveness”) were more concentrated in densely-populated areas in 2000, but not in 1880. We provide evidence in support of the predictions of a model of agglomeration and the task composition of employment, in which reductions in trade costs lead to an increased concentration of interactive tasks in metro areas.

**KEYWORDS:** Agglomeration, Task Specialization, Urbanization.

**JEL CLASSIFICATION:** N92, O18, R12

---

\*We are grateful to the Centre for Economic Performance and Princeton University for research support. We are also grateful to Gilles Duranton and Matt Turner for sharing data. We thank William D. Caughlin and George Kupczak for their help with data. We thank Don Davis, Gilles Duranton, Gene Grossman, Jessie Handbury, Gordon Hanson, Vernon Henderson, Rick Hornbeck, Jeffrey Lin, Marco Manacorda, Alan Manning, Nathan Nunn, Gianmarco Ottaviano, Steve Pischke, Nancy Qian, Esteban Rossi-Hansberg, Yona Rubinstein, Will Strange, Tony Venables and seminar and conference participants at CEPR, Copenhagen, CURE, the London School of Economics, MIT, NARSC, Oxford, Sorbonne, Sussex and UCSD for helpful comments and suggestions. The usual disclaimer applies.

<sup>†</sup>CEP, LSE, Houghton Street, WC2A 2AE, London, UK. Email: g.michaels@lse.ac.uk. Tel: +44(0)20-7852-3518.

<sup>‡</sup>Department of Economics, University of Oxford, Manor Road, Oxford, OX1 3UQ. Email: ferdiand.rauch@economics.ox.ac.uk. Tel: +44(0)77214-88037.

<sup>§</sup>Fisher Hall, Princeton, NJ 08540. Email: reddings@princeton.edu. Tel: +1(609) 258-4016.

# 1 Introduction

As a large and growing share of the world’s population concentrates in cities, workers’ labor market outcomes are increasingly determined in urban labor markets. Yet we know relatively little about the tasks undertaken in urban versus rural areas and how these have changed over time. This paper combines Census of Population and Dictionary of Occupations (DOTs) data to provide novel evidence on the task composition of employment in urban and rural areas in the United States from 1880-2000. What kinds of jobs are available in urban areas? How do these differ from the jobs available in rural areas? Do large and small cities offer the same employment opportunities or does the task composition of employment differ across cities of different sizes? Are these patterns stable over time? If they have changed, what explains these changes in patterns of task specialization? What are the implications of the evolving task composition of employment for our understanding of the sources of agglomeration? This paper provides answers to these and other related questions.

We use the DOTs data to provide the first evidence on the evolution of the task composition of employment in urban and rural areas over a long historical time period. To measure the tasks undertaken by workers within occupations, we examine the occurrence of 3,000 verbs in over 12,000 occupational descriptions. We find substantial changes in the tasks undertaken within urban versus rural areas. In 1880, the tasks most concentrated in metro areas included “Braid,” “Sew,” “Stretch” and “Thread.” In contrast, in 2000, the tasks most concentrated in metro areas comprised “Analyze,” “Advise,” “Confer” and “Report.” We use Roget’s Thesaurus – the seminal reference for English language use – to interpret these changing patterns of task specialization. We find a negative and statistically significant correlation of -0.43 between the thesaurus categories most concentrated in metro areas in 1880 and 2000. In particular, we find an increase over time in the concentration in metro areas of tasks involving a combination of thought, communication and intersocial activity. We define this combination of tasks as “interactiveness,” because it captures the essence of human interaction: the generation (“thought”) and transmission (“communication”) of ideas to other humans (“intersocial activity”). At the beginning of our sample period, metro and non-metro areas actually had similar levels of interactiveness. Over time, metro areas experienced larger increases in interactiveness than non-metro areas, making metro areas substantially more interactive today. We date the origins of this rise in the relative interactiveness of metro areas back to the early-twentieth century.

Understanding changes in task specialization in urban and rural areas is central to evaluating alternative theories of agglomeration. Traditionally, these theories have emphasized the costs of moving goods and people over space. In the new economic geography literature, agglomeration is explained by consumer love of differentiated varieties, increasing returns to scale and the costs of transporting these differentiated varieties (e.g. Fujita, Krugman and Venables 1999). In the canonical models of urban economies, agglomeration is driven by economies of scale in producing a single final good, while commuting costs provide a

dispersion force (e.g. Alonso 1964, Muth 1968, Mills 1967). More recently, research has focused on human capital externalities (e.g. Moretti 2004, Davis and Dingel 2013) and the costs of exchanging ideas (e.g. Davis and Dingel 2012, Gaspar and Glaeser 1998), which highlights the costs of moving ideas rather than goods and people (e.g. Glaeser and Kohlhase 2003).

Although these theories emphasize quite different sources of agglomeration economies, there is relatively little empirical evidence on the relative importance of alternative sources of agglomeration and how this has changed over time. To shed further light on these questions, we measure the task content of employment and its evolution in urban and rural areas over time. Our findings of an increased importance of interactive tasks in urban relative to rural areas support the idea that the movement of ideas has become more important as a source of agglomeration relative to the movement of goods and people.

Since externalities in the movement of ideas are likely to differ from those in the movement of goods and people, our findings suggest potential changes in the strength of agglomeration economies over time, with implications for the alignment of private and social returns to location choices. Whereas most existing empirical research has concentrated on estimating the overall strength of agglomeration economies, our findings highlight the need to distinguish between different sources of agglomeration economies and allow their relative importance to change over time.<sup>1</sup> Our results suggest that rapid urbanization is likely to be accompanied by substantial changes in the relative demand for skills across occupations and tasks.

To guide our empirical analysis, we develop a model that captures specialization across the many sectors, occupations and locations observed in our data. We use the model to rationalize a secular reallocation of employment towards interactive occupations across all locations and a greater reallocation of employment towards interactive occupations in more densely-populated locations. The distribution of economic activity across locations in the model is determined by productivity differences (including agglomeration forces which concentrate population in productive locations) and an inelastic supply of land (which favors population dispersion). Each location can produce in a number of sectors (e.g. Manufacturing, Services, which correspond to the two-digit sectors in our data) and each sector includes a number of more disaggregated goods (e.g. Motor Vehicles, Drugs and Medicines, which correspond to the three-digit sectors in our data). Production of each good involves a number of stages of production supplied by occupations (e.g. Managers, Operatives, which correspond to the two-digit occupations in our data) and workers within in each occupation perform a number of tasks (e.g. advising, typing, stamping, stretching, which correspond to the verbs in the descriptions for the disaggregated occupations in our data). Final goods can be traded between locations subject to goods trade costs that can differ across sectors. Tasks also can be traded between locations subject to task trade costs that differ across occupations.

As in the macroeconomics literature on structural transformation, the model can account for a secu-

---

<sup>1</sup>Few empirical studies have sought to distinguish between different sources of agglomeration economies, a notable exception being Ellison, Glaeser and Kerr (2010).

lar reallocation of employment across sectors and occupation in all locations. When demand is inelastic for sectors and occupations, more rapid productivity growth in some sectors and occupations leads to a more than proportionate fall in price, which reallocates employment towards other sectors and occupations. As in the international trade literature on offshoring, the model can account for a greater reallocation of employment towards interactive locations in more densely-populated locations through improvements in communications and transport technologies. Reductions in final goods trade costs induce specialization across sectors according to standard theories of comparative advantage. Reductions in task trade costs induce an analogous process of specialization across occupations within sectors. When task trade costs are prohibitively high, all tasks are performed in the location in which the final good is produced. As task trade costs fall, it becomes feasible to unbundle production across locations and trade tasks between these locations. To the extent that densely populated locations have a comparative advantage in interactive tasks (e.g. because agglomeration forces are stronger for interactive tasks), reductions in task trade costs induce densely-populated locations to specialize in more-interactive occupations, while more sparsely-populated locations specialize in less-interactive occupations.

Guided by these predictions, we develop a methodology for measuring the tasks undertaken within each occupation, and provide evidence on the evolution of these tasks in urban and rural areas over time. We show that densely-populated locations have become more interactive relative to less-densely-populated locations – both between metro and non-metro areas and across metro areas of different population densities. We find similar results for each of the components of interactiveness (thought, communication and intersocial activity), suggesting that it is the combination of these activities that is important. We provide external validation for our results using separate independent measures of interactiveness that have been constructed for contemporary time periods. We establish that the concentration of interactive occupations in metro areas is observed for both employment and wage bill shares, consistent with an increase in the relative demand for interactive occupations.

We also provide evidence against a number of potential alternative explanations. To demonstrate that our findings are not driven by the relocation of manufacturing from urban areas or the expansion of services in urban areas, we show that the increased interactiveness of metro areas takes place within as well as between sectors and is not driven by any one occupation or sector. We also find similar results when we exclude occupations associated with headquarters, suggesting that the increased interactiveness is not driven solely by a concentration of headquarters in urban areas. We show that our findings are not explained by an expansion in the geographical boundaries of metro areas over time, since we find similar results when we restrict attention to central cities. We document similar results for single and married individuals, suggesting that the increased interactiveness of metro areas is not driven by an increased need for “power couples” to collocate. Taken together, these results support a change in the organization of production activities within industries to favor the concentration of more interactive tasks in cities.

Our paper is related to a number of literatures. We build on the wider literature on agglomeration economies, as surveyed in Duranton and Puga (2004) and Rosenthal and Strange (2004). One strand of this literature emphasizes the role of human capital, skills and the division of labor in promoting agglomeration, including Baumgardner (1988), Becker and Murphy (1992), Glaeser and Saiz (2003), Glaeser and Resseger (2009), Bacolod, Blum and Strange (2009a,b), Duranton (1998), Duranton and Jayet (2011), Glaeser, Ponzetto and Toblo (2011), Henderson (1974), Lin (2011) and Moretti (2004). Another strand of this literature has contrasted specialization by sector with specialization by function (e.g. headquarters versus plants), including Duranton and Puga (2005), Rossi-Hansberg, Sarte and Owens (2009), Ota and Fujita (1993), Helsley and Strange (2007), and Fujita and Tabuchi (1997). In contrast to these studies, which have used relatively aggregate measures of skills and specialization, we use the disaggregated information in the DOTs to provide the first evidence on tasks undertaken in metro and non-metro areas over time.

Our research also builds on the labor economics literature on the task content of employment, including Autor, Levy and Murnane (2003), Autor and Dorn (2013), Autor and Handel (2009), and Gray (2010). We develop a new approach to measuring the task content of employment, which enables us to measure the individual tasks performed by workers within each occupation (captured by the verbs in occupational descriptions) rather than being restricted to a single overall numerical score for each occupation (such as “Direction, Control and Planning (DCP)”). We use these verbs to measure the relative importance of multiple tasks for each occupation and to develop a new measure of the overall interactivensness of occupations. We thus contribute to an emerging literature in economics and the social sciences that uses textual search to quantify hard-to-measure concepts, including political influence in Gentzkow et al. (2012) and culture in Michel et al. (2011). In contrast to much of the labor economics literature, our focus is on differences in the task content of employment between urban and rural areas. Whereas this prior labor economics research has mainly concentrated on recent decades, we provide evidence over more than a century.

Our theoretical and empirical research on the remote sourcing of tasks within countries is also related to the international trade literature on offshoring, including Becker, Ekholm and Muendler (2009), Blinder (2009), Blinder and Krueger (2012), Grossman and Rossi-Hansberg (2008), and Ottaviano, Peri and Wright (2010). Our results suggest that the remote sourcing of tasks can also be between locations within countries.

The remainder of the paper is structured as follows. Section 2 introduces the model. Section 3 discusses the data. Section 4 presents some initial evidence based on sectors and occupations to motivate our more detailed analysis of task specialization. Section 5 introduces our empirical methodology for measuring task specialization. Section 6 constructs our measure of interactivensness and presents our baseline results. Section 7 reports a number of robustness tests and compares our measure of an occupation’s interactivensness to other independent measures and to other occupational characteristics. Section 8 presents evidence on the determinants of changes in task specialization. Section 9 concludes.

## 2 Theoretical Model

In this section, we outline a theoretical model that we use to rationalize the empirical results reported below.<sup>2</sup> First, the model explains a secular reallocation of employment over time towards more interactive occupations in terms of differences in productivity growth across occupations and inelastic demand. Second, the model predicts that this reallocation is stronger in metro areas than in non-metro areas because of falling trade costs and specialization according to comparative advantage. To the extent that densely-populated locations are relatively more productive in interactive tasks (because agglomeration forces are stronger for interactive tasks), reductions in task trade costs induce densely-populated locations to specialize in more-interactive occupations, while more sparsely-populated locations specialize in less-interactive occupations. Third, and for the same reason, the model predicts that the reallocation towards more interactive occupations is stronger in denser metro areas. Fourth, the model explains how this allocation of tasks across areas that differ in their population density could have taken place both between sectors and within sectors. In this respect our model is more general than existing models that focus on differential changes across sectors (e.g. manufacturing vs. services) or within sectors (e.g. relocation of headquarters). Fifth, the model accounts for our finding that in 1880, when trade in tasks was still costly, metro areas were not more interactive than non-metro areas. Finally, the model explains faster population growth over our sample period in metro than in non-metro areas in terms of faster productivity growth in metro areas.

While some existing models may be consistent with some of these stylized facts, we are unaware of any model that can account for all of them. One of the challenges in developing such a model is retaining tractability despite allowing for rich patterns of specialization across occupations, sectors and locations. To ensure that the model remains tractable, we use a stochastic formulation of productivity differences across occupations, sectors and locations. The distribution of population across locations is determined as the outcome of a tension between congestion forces (an inelastic supply of land) and agglomeration forces (productivity differences that depend on production externalities). Variation in the magnitude of these productivity differences across occupations and sectors gives rise to specialization according to Ricardian comparative advantage within and across sectors.

### 2.1 Preferences and Endowments

The economy consists of many locations indexed by  $n \in N$ . Each location  $n$  is endowed with an exogenous supply of land  $\bar{H}_n$ . The economy as a whole is endowed with a measure of workers  $\bar{L}$ , who are perfectly mobile across locations. Workers' preferences are defined over a goods consumption index ( $C_n$ )

---

<sup>2</sup>A more detailed exposition of the model including the technical derivations of relationships is contained in a web-based technical appendix.

and residential land use ( $H_n$ ) and are assumed to take the Cobb-Douglas form:<sup>3</sup>

$$U_n = C_n^\alpha H_n^{1-\alpha}, \quad 0 < \alpha < 1. \quad (1)$$

The goods consumption index ( $C_n$ ) is assumed to be a constant elasticity of substitution (CES) function of consumption indices for a number of sectors (e.g. Manufacturing, Services) indexed by  $s \in S$ :

$$C_n = \left[ \sum_{s \in S} C_{ns}^{\frac{\beta-1}{\beta}} \right]^{\frac{\beta}{\beta-1}}, \quad (2)$$

where  $\beta$  is the elasticity of substitution between goods. Sectors can be either substitutes ( $\beta > 1$ ) or complements in goods consumption ( $0 < \beta < 1$ ), where the standard assumption in the literature on structural transformation in macroeconomics is complements (e.g. Ngai and Pissarides 2007, Yi and Zhang 2010).

The consumption index for each sector is in turn a CES function of consumption of a continuum of goods (e.g. Motor Vehicles, Drugs and Medicines) indexed by  $j \in [0, 1]$ :

$$C_{ns} = \left[ \int_0^1 c_{ns}(j)^{\frac{\sigma_s-1}{\sigma_s}} dj \right]^{\frac{\sigma_s}{\sigma_s-1}}, \quad (3)$$

where the elasticity of substitution between goods  $\sigma_s$  varies across sectors. While in the data we observe a finite number of goods within sectors, we adopt the theoretical assumption of a continuum of goods within sectors for reasons of tractability, because it enables us to make use of law of large numbers results in determining specialization at the sectoral level. Goods can be either substitutes ( $\sigma_s > 1$ ) or complements ( $0 < \sigma_s < 1$ ) and we can allow any ranking of the elasticities of substitution between goods and sectors, although the conventional assumption in such a nested CES structure is a higher elasticity of substitution at the more disaggregated level ( $\sigma_s > \beta$ ).

Expenditure on residential land in each location is assumed to be redistributed lump-sum to residents of that location, as in Helpman (1998). Therefore total income in each location equals payments to labor used in production plus expenditure on residential land:

$$v_n L_n = w_n L_n + (1 - \alpha) v_n L_n = \frac{w_n L_n}{\alpha}, \quad (4)$$

where  $w_n$  is the wage;  $L_n$  is the population of location  $n$ ; and equilibrium land rents in each location ( $r_n$ ) are determined by land market clearing.

## 2.2 Production

Goods are homogeneous in the sense that one unit of a given good is the same as any other unit of that good. Production occurs under conditions of perfect competition and constant returns to scale. The cost to

---

<sup>3</sup>For empirical evidence using U.S. data in support of the constant expenditure share implied by the Cobb-Douglas functional form, see Davis and Ortalo-Magne (2011).

a consumer in location  $n$  of purchasing one unit of good  $j$  within sector  $s$  from location  $i$  is therefore:

$$p_{nis}(j) = \frac{d_{nis}G_{is}(j)}{z_{is}(j)}, \quad (5)$$

where  $d_{nis}$  are iceberg goods trade costs, such that  $d_{nis} > 1$  must be shipped from location  $i$  to location  $n$  within sector  $s$  in order for one unit to arrive;  $z_{is}(j)$  is productivity for good  $j$  within sector  $s$  in location  $i$ ; and  $G_{is}(j)$  is the unit cost of the composite factor of production used for good  $j$  within sector  $s$  in location  $i$ , as determined below.

Final goods productivity is stochastic and modeled as in Eaton and Kortum (2002) and Costinot, Donaldson and Komunjer (2012). Final goods productivity for each good, sector and location is assumed to be drawn independently from a Fréchet distribution:<sup>4</sup>

$$F_{is}(z) = e^{-T_{is}L_{is}^{\eta_s}z^{\theta_s}}, \quad (6)$$

where the shape parameter ( $\theta_s > 1$ ) controls the dispersion of productivity across goods within each sector, which determines comparative advantage across goods. In contrast, the scale parameter ( $T_{is}L_{is}^{\eta_s}$ , where  $\eta_s > 0$ ) determines average productivity within each sector for each location, which in turn determines comparative advantage across sectors. We allow average productivity in a sector and location to be increasing in employment in that sector and location to capture agglomeration forces in the form of external economies of scale in final goods production (e.g. Ethier 1982). As these external economies are external to the firm, they are consistent with our assumption of perfect competition, since each firm takes productivity as given when making its decisions.

As in the Ricardian model of trade, our framework features comparative advantage across goods and sectors, which explains the observed variation in the shares of sectors and goods in employment across locations in the data. But we also observe variation in the shares of occupations and tasks in employment across locations in the data, which is not captured in the standard Ricardian framework. To account for this additional layer of specialization, we assume that each good is produced using a number of stages of production, where each stage of production within a sector is supplied by a separate occupation indexed by  $o \in O_s$  (e.g. Managers, Operatives). Output of good  $j$  within sector  $s$  in location  $i$  ( $y_{is}(j)$ ) is a CES function of the inputs of each occupation ( $X_{iso}(j)$ ):

$$y_{is}(j) = \left[ \sum_{o \in O_s} X_{iso}(j)^{\frac{\mu_s - 1}{\mu_s}} \right]^{\frac{\mu_s}{\mu_s - 1}}, \quad (7)$$

where  $\mu_s$  is the elasticity of substitution between occupations and again we can allow occupations to be either substitutes ( $\mu_s > 1$ ) or complements ( $0 < \mu_s < 1$ ). We allow sectors to differ in terms of the

<sup>4</sup>To simplify the exposition, we use  $i$  to denote locations of production and  $n$  to denote locations of consumption, except where otherwise indicated.

set of occupations  $O_s$ , and firms within each sector adjust the proportions with which workers in different occupations are employed depending their cost.

Workers within each occupation perform a continuum of tasks  $t \in [0, 1]$  as in Grossman and Rossi-Hansberg (2008) (e.g. Advising, Typing, Stretching, Stamping). The input for occupation  $o$  and good  $j$  within sector  $s$  and location  $i$  ( $X_{iso}(j)$ ) is a CES function of the inputs for these tasks ( $x_{iso}(j, t)$ ):

$$X_{iso}(j) = \left[ \int_0^1 x_{iso}(j, t)^{\frac{\nu_{so}-1}{\nu_{so}}} dt \right]^{\frac{\nu_{so}}{\nu_{so}-1}}, \quad (8)$$

where the elasticity of substitution between tasks  $\nu_{so}$  varies across sectors and occupations. While in the data we observe a finite number of tasks within occupations, we adopt the theoretical assumption of a continuum of tasks within occupations for reasons of tractability, because it enables us to make use of law of large numbers results in determining specialization at the occupational level.<sup>5</sup> We allow tasks within occupations to be either substitutes ( $\nu_{so} > 1$ ) or complements ( $0 < \nu_{so} < 1$ ), and we can consider any ranking of the elasticities of substitution between tasks and occupations, although the conventional assumption in such a nested CES structure is again a higher elasticity of substitution at the more disaggregated level ( $\nu_{so} > \mu_s$ ).<sup>6</sup>

Tasks are performed by labor using a constant returns to scale technology and can be traded between locations. For example, product design can be undertaken in one location, while production and assembly occur in another location. The cost to a firm in location  $n$  of sourcing a task  $t$  from location  $i$  within occupation  $o$  and sector  $s$  is:

$$g_{niso}(j, t) = \frac{\tau_{niso} w_i}{a_{iso}(j, t)}, \quad (9)$$

where  $w_i$  is the wage;  $\tau_{niso}$  are iceberg task trade costs, such that  $\tau_{niso} > 1$  units of the task must be performed in location  $i$  in order for one unit to be completed in location  $n$  for occupation  $o$  and sector  $s$ ;  $a_{iso}(j, t)$  is productivity for task  $t$  and good  $j$  within occupation  $o$  and sector  $s$  in location  $i$ .

Input productivity for each task, occupation, sector and location is also stochastic and is assumed to be drawn independently from a Fréchet distribution:

$$\mathcal{F}_{iso} = e^{-U_{iso} L_{iso}^{\chi_{so}} a^{-\epsilon_{so}}}, \quad (10)$$

where the shape parameter ( $\epsilon_{so} > 1$ ) controls the dispersion of productivity across tasks within occupations, which determines comparative advantage across tasks. In contrast, the scale parameter ( $U_{iso} L_{iso}^{\chi_{so}} > 0$ , where  $\chi_{so} > 0$ ) controls average productivity within each occupation, which determines comparative advantage across occupations. We allow average productivity in an occupation, sector and location to be increasing in employment in that occupation, sector and location ( $\chi_{so} > 0$ ) to capture external economies of scale in

<sup>5</sup>To reduce the notational burden, we assume the same  $[0, 1]$  interval of tasks for all occupations, but it is straightforward to allow this interval to vary across occupations.

<sup>6</sup>While we interpret production as being undertaken by workers in occupations that perform many tasks, an equivalent interpretation is that each occupation corresponds to a stage of production and each task corresponds to an intermediate input within that stage of production.

task production (e.g. Grossman and Rossi-Hansberg 2012). As these increasing returns to scale are again external to the firm, they are consistent with our assumption of perfect competition, since each firm takes productivity as given when making its decisions.

### 2.3 Trade in Tasks and Input Costs

Firms within a given location  $n$  source each task  $t$  within an occupation  $o$ , good  $j$  and sector  $s$  from the lowest cost source of supply for that task:

$$g_{nso}(j, t) = \min \{g_{niso}(j, t); i \in N\}.$$

Given finite task trade costs, locations supply tasks for which they have high productivity draws themselves, and source other tasks for which they have low productivity draws from other locations. Under our assumption of a Fréchet distribution of input productivity, the share of firm costs in location  $n$  accounted for by tasks sourced from location  $i$  within occupation  $o$  and sector  $s$  ( $\lambda_{niso}$ ) is equal to the fraction of tasks sourced from that location:

$$\lambda_{niso} = \frac{U_{iso} L_{iso}^{\chi_{so}} (\tau_{niso} w_i)^{-\epsilon_{so}}}{\sum_{k \in N} U_{kso} L_{kso}^{\chi_{so}} (\tau_{nkso} w_k)^{-\epsilon_{so}}}, \quad (11)$$

and the unit cost for occupation  $o$  and sector  $s$  in location  $n$  can be written as:

$$G_{nso} = \gamma_{so} \left( \frac{U_{nso} L_{nso}^{\chi_{so}}}{\lambda_{nnso}} \right)^{-\frac{1}{\epsilon_{so}}} w_n. \quad (12)$$

where  $\gamma_{so} = \left[ \Gamma \left( \frac{\epsilon_{so} + 1 - \nu_{so}}{\epsilon_{so}} \right) \right]^{\frac{1}{1 - \nu_{so}}}$  and  $\Gamma(\cdot)$  is the gamma function.

Intuitively, costs for occupation  $o$  in sector  $s$  and location  $n$  are low when the average input productivity for performing tasks with that occupation, sector and location ( $U_{nso} L_{nso}^{\chi_{so}}$ ) is high, when the wage ( $w_n$ ) is low, and when that occupation, sector and location spends a low share of its costs on itself ( $\lambda_{nnso}$ ). This low own trade share in costs reflects the ability of the occupation, sector and location to lower its costs by remotely-sourcing tasks from other locations, and hence trade in tasks acts like an increase in productivity that reduces unit costs.

Using unit costs for each occupation (11) and the final goods production technology (7), the share of occupation  $o$  in sector costs is:

$$e_{nso} = \frac{\gamma_{so}^{1 - \mu_s} \left( \frac{U_{nso} L_{nso}^{\chi_{so}}}{\lambda_{nnso}} \right)^{-\frac{1 - \mu_s}{\epsilon_{so}}}}{\sum_{m \in O_s} \gamma_{sm}^{1 - \mu_s} \left( \frac{U_{nsm} L_{nsm}^{\chi_{so}}}{\lambda_{nsm}} \right)^{-\frac{1 - \mu_s}{\epsilon_{sm}}}}. \quad (13)$$

Intuitively, high-unit-cost occupations (low  $U_{nso} L_{nso}^{\chi_{so}} / \lambda_{nnso}$ ) account for high shares of firm costs and employment if occupations are complements ( $0 < \mu_s < 1$ ) and low shares of firm costs and employment if occupations are substitutes ( $\mu_s > 1$ ).

## 2.4 Trade in Final Goods and Price Indices

Consumers within a given location  $n$  source each final good  $j$  within a sector  $s$  from the lowest cost source of supply for that final good:

$$p_{ns}(j) = \min \{p_{nis}(j); i \in N\}.$$

For finite final goods trade costs, locations supply final goods for which they have low unit costs themselves, and source other final goods for which they have high units costs from other locations. These unit costs for final goods depend on input productivities and trade in tasks, as characterized in the previous subsection, as well on final goods productivities. Under our assumption of a Fréchet distribution of final goods productivity, the share of location  $n$ 's expenditure on final goods produced in location  $i$  within sector  $s$  ( $\pi_{nis}$ ) is equal to the fraction of final goods sourced from that location:

$$\pi_{nis} = \frac{T_{is} L_{is}^{\eta_s} (d_{nis} \Phi_{is} w_i)^{-\theta_s}}{\sum_{k \in N} T_{ks} L_{ks}^{\eta_s} (d_{nks} \Phi_{ks} w_k)^{-\theta_s}}, \quad (14)$$

and the price index for sector  $s$  in location  $n$  can be written as:

$$P_{ns} = \kappa_s \left( \frac{T_{ns} L_{ns}^{\eta_s}}{\pi_{nns}} \right)^{-\frac{1}{\theta}} \Phi_{ns} w_n. \quad (15)$$

where  $\Phi_{is}$  is a summary statistic for occupation unit costs in sector  $s$  in location  $i$ :

$$\Phi_{is} = \left[ \sum_{o \in O_s} \gamma_{so}^{1-\mu_s} \left( \frac{U_{iso} L_{iso}^{\chi_{so}}}{\lambda_{iiso}} \right)^{-\frac{1-\mu_s}{\epsilon_{so}}} \right]^{\frac{1}{1-\mu_s}}. \quad (16)$$

and  $\kappa_s = \left[ \Gamma \left( \frac{\theta_s + 1 - \sigma_s}{\theta_s} \right) \right]^{\frac{1}{1-\sigma_s}}$  and  $\Gamma(\cdot)$  is the Gamma function.

Intuitively, the price index for sector  $s$  in location  $n$  is low when average productivity within that sector and location ( $T_{ns} L_{ns}^{\eta_s}$ ) is high, when the own trade share for final goods within that sector and location ( $\pi_{nns}$ ) is low, and when unit costs for that sector and location are low ( $\Phi_{ns} w_n$ ). Unit costs in turn can be low because of a low wage ( $w_n$ ), high average input productivities for occupations within that sector and location ( $U_{nso} L_{nso}^{\chi_{so}}$ ), and low own trade shares for occupations within that sector and location ( $\lambda_{nso}$ ).

Using the sectoral price index (15), the share of sector  $s$  in aggregate goods expenditure is:

$$E_{ns} = \frac{\kappa_s^{1-\beta} \left( \frac{T_{ns} L_{ns}^{\eta_s}}{\pi_{nns}} \right)^{-\frac{1-\beta}{\theta_s}} \Phi_{ns}^{1-\beta}}{\sum_{r \in S} \kappa_r^{1-\beta} \left( \frac{T_{nr} L_{nr}^{\eta_r}}{\pi_{nnr}} \right)^{-\frac{1-\beta}{\theta_r}} \Phi_{nr}^{1-\beta}}. \quad (17)$$

Intuitively, high-price sectors (low  $T_{ns} L_{ns}^{\eta_s}$ , high  $\pi_{nns}$  and high  $\Phi_{ns}$ ) account for high shares of aggregate goods expenditure and employment if sectors are complements ( $0 < \beta < 1$ ) and low shares of aggregate goods expenditure and employment if sectors are substitutes ( $\beta > 1$ ).

## 2.5 Population Mobility

Population mobility implies that workers receive the same indirect utility in all populated locations:

$$V_n = \frac{v_n}{P_n^\alpha r_n^{1-\alpha}} = \bar{V}, \quad (18)$$

while labor market clearing requires:

$$\sum_{n \in N} L_n = \bar{L}. \quad (19)$$

In the web appendix, we characterize the general equilibrium of the model, including the distribution of population across locations ( $L_n$ ) and each location's own trade share for final goods ( $\pi_{nns}$ ) and tasks ( $\lambda_{nns}$ ). The stochastic formulation of productivity for both goods and tasks ensures that the model remains tractable and yields deterministic predictions for patterns of specialization across both sectors and occupations. In the next two subsections, we use these predictions to guide our empirical analysis.

## 2.6 Secular Reallocation towards Interactive Occupations

One key feature of our data is a reallocation of employment towards more interactive occupations over time, which occurs across all locations and both between and within sectors. As in the macroeconomics literature on structural transformation, the model accounts for such secular reallocation in terms of differences in productivity growth across sectors and inelastic demand between sectors (e.g. Ngai and Pissarides 2007, Yi and Zhang 2010). To show this formally, partition final goods productivity in a sector-location into a sector component ( $\tilde{T}_s$ ), a location component ( $\tilde{T}_n$ ) and a residual ( $\tilde{T}_{ns}$ ):  $T_{ns} = \tilde{T}_s \tilde{T}_n \tilde{T}_{ns}$ . Since the sector component  $\tilde{T}_s$  is common to all locations, it cancels from the numerator and denominator of the location trade share ( $\pi_{nis}$ ) and hence does not directly effect  $\pi_{nis}$  for given wages. In contrast, the sector component  $\tilde{T}_s$  directly affects the share of sectors in aggregate expenditure ( $E_{ns}$ ) for all locations. Taking the partial derivative of this expenditure share (17) with respect to  $\tilde{T}_s$  at the initial equilibrium vectors of wages ( $\mathbf{w}$ ) and employment ( $\mathbf{L}_{so}$ ), faster productivity growth in sector  $s$  reduces the share of that sector in expenditure and increases the share of all other sectors in expenditure when sectors are complements and has the reverse effect when sectors are substitutes:

$$\begin{aligned} \left. \frac{\partial E_{ns}}{\partial \tilde{T}_s} \frac{\tilde{T}_s}{E_{ns}} \right|_{\mathbf{w}, \mathbf{L}_{so}} &= - \left( \frac{1-\beta}{\theta_s} \right) (1 - E_{ns}) < 0, & 0 < \beta < 1, \\ \left. \frac{\partial E_{nr}}{\partial \tilde{T}_s} \frac{\tilde{T}_s}{E_{nr}} \right|_{\mathbf{w}, \mathbf{L}_{so}} &= \left( \frac{1-\beta}{\theta_s} \right) E_{nr} > 0, & r \neq s, 0 < \beta < 1, \end{aligned}$$

where this productivity growth and the changes in expenditure shares it induces in turn have general equilibrium effects for wages and employment allocations.

While the macroeconomics literature on structural transformation typically focuses on sectors, secular changes in the shares of occupations in employment across all locations can be explained in the model by an analogous process of differences in productivity growth across occupations and inelastic demand between

occupations. To show this formally, partition average input productivity in an occupation-sector-location into an occupation component ( $\tilde{U}_o$ ), sector component ( $\tilde{U}_s$ ), location component ( $\tilde{U}_n$ ) and a residual ( $\tilde{U}_{nso}$ ):  $U_{nso} = \tilde{U}_o \tilde{U}_s \tilde{U}_n \tilde{U}_{nso}$ . Since the occupation component  $\tilde{U}_o$  is common to all locations, it cancels from the numerator and denominator of the location trade share ( $\lambda_{niso}$ ) and hence does not directly effect  $\lambda_{niso}$  for given wages. In contrast, the occupation component  $\tilde{U}_o$  directly affects the share of occupations in sectoral expenditure ( $e_{nso}$ ) for all locations. Taking the partial derivative of this expenditure share (13) with respect to  $\tilde{U}_o$  at the initial equilibrium vectors of wages ( $\mathbf{w}$ ) and employment ( $\mathbf{L}_{so}$ ), faster productivity growth in occupation  $o$  reduces the share of that occupation in costs and increases the share of all other occupations in costs when occupations are complements and has the reverse effect when occupations are substitutes:

$$\begin{aligned} \left. \frac{\partial e_{nso}}{\partial \tilde{U}_o} \frac{\tilde{U}_o}{e_{nso}} \right|_{\mathbf{w}, \mathbf{L}_{so}} &= - \left( \frac{1-\mu_s}{\epsilon_{so}} \right) (1 - e_{nso}) < 0, & 0 < \mu_s < 1, \\ \left. \frac{\partial e_{nsm}}{\partial \tilde{U}_o} \frac{\tilde{U}_o}{e_{nsm}} \right|_{\mathbf{w}, \mathbf{L}_{so}} &= \left( \frac{1-\mu_s}{\epsilon_{so}} \right) e_{nsm} > 0, & m \neq o, 0 < \mu_s < 1, \end{aligned}$$

where this productivity growth and the changes in expenditure shares it induces in turn have general equilibrium effects for wages and employment allocations.

## 2.7 Increased Relative Interactiveness of More Densely-populated Locations

Another key feature of our data is that the reallocation of employment towards more interactive occupations over time is more pronounced in more densely-populated locations. The model explains these differential changes in employment across occupations by specialization according to comparative advantage.

This specialization occurs across occupations within sectors and can be characterized by a double difference across exporting locations and occupations within an importing location. The first difference computes the ratio of exports of tasks from two locations  $i$  and  $k$  in a third market  $n$  in a single occupation; the second difference compares this ratio of exports of tasks for two separate occupations  $o$  and  $m$ :

$$\frac{\lambda_{niso}/\lambda_{nkso}}{\lambda_{nism}/\lambda_{nksm}} = \frac{[U_{iso} L_{iso}^{\chi_{so}} (\tau_{niso} w_i)^{-\epsilon_{so}}] / [U_{kso} L_{kso}^{\chi_{so}} (\tau_{nkso} w_k)^{-\epsilon_{so}}]}{[U_{ism} L_{ism}^{\chi_{so}} (\tau_{nism} w_i)^{-\epsilon_{sm}}] / [U_{ksm} L_{ksm}^{\chi_{so}} (\tau_{nksm} w_k)^{-\epsilon_{sm}}]}.$$

From the above double difference, locations export relatively more tasks in occupations in which they have relatively lower costs of supply, where these costs of supply depend on relative productivities (which in turn depend on relative employments through the external economies of scale), relative wages and task trade costs. Each location is a net exporter of tasks in some occupations and a net importer of tasks in other occupations (inter-occupation trade in tasks). This inter-occupation trade generates differences across locations in the pattern of employment across occupations within sectors.

There is also specialization according to comparative advantage across sectors, which can be characterized by an analogous double difference across exporting locations and sectors within an importing location. The first difference computes the ratio of exports of final goods from two locations  $i$  and  $k$  in a third market

$n$  in a single sector; the second difference compares this ratio of exports of final goods for two separate sectors  $s$  and  $r$ :

$$\frac{\pi_{nis}/\pi_{nks}}{\pi_{nir}/\pi_{nkr}} = \frac{\left[ T_{is} L_{is}^{\eta_s} (d_{nis} \Phi_{is} w_i)^{-\theta_s} \right] / \left[ T_{ks} L_{ks}^{\eta_s} (d_{nks} \Phi_{ks} w_k)^{-\theta_s} \right]}{\left[ T_{ir} L_{ir}^{\eta_r} (d_{nir} \Phi_{ir} w_i)^{-\theta_r} \right] / \left[ T_{kr} L_{kr}^{\eta_r} (d_{nkr} \Phi_{kr} w_k)^{-\theta_r} \right]}.$$

From the above double difference, locations export relatively more final goods in sectors in which they have relatively lower costs of supply, where these costs of supply depend on relative productivities (which in turn depend on relative employments through the external economies of scale and on trade in tasks), relative unit costs (which depend on wages and trade in tasks), and final goods trade costs. Each location is a net exporter of final goods in some sectors and a net importer of final goods in other sectors (inter-industry trade in goods). This inter-industry trade generates differences across locations in the pattern of employment across sectors.

Reductions in final goods trade costs ( $d_{nis}$ ) induce specialization across sectors according to standard theories of comparative advantage. Reductions in task trade costs ( $\tau_{niso}$ ) induce an analogous process of specialization across occupations within sectors. When task trade costs are prohibitively high, all tasks are performed in the location in which the final good is produced. As task trade costs fall, it becomes feasible to unbundle production across locations and trade tasks between these locations.<sup>7</sup> To the extent that densely-populated locations are relatively more productive in interactive tasks (e.g. because agglomeration forces  $\chi_{so}$  are stronger for interactive tasks), reductions in task trade costs induce densely-populated locations to specialize in more-interactive occupations, while more sparsely-populated locations specialize in less-interactive occupations.<sup>8</sup> Thus the model accounts for an increase in the interactiveness of more-densely-populated locations relative to less-densely-populated locations during our sample period in terms of falling trade costs and increased specialization according to comparative advantage.

### 3 Data Description

Our empirical analysis uses two main sources of data. The first is individual-level records from the U.S. Population Census for twenty-year intervals from 1880-2000 from Integrated Public Use Microdata Series (IPUMS): see Ruggles et al. (2010). These census micro data report individuals' location, occupation and sector, as well as other demographic information. We use these data to determine whether an individual is located in a metro area as well as the occupation and sector in which an individual is employed.<sup>9</sup> We weight

<sup>7</sup>For further discussion of the increased unbundling of production, see for example Baldwin (2012).

<sup>8</sup>While our model focuses on trade in tasks across locations within countries, a similar process of trade in tasks could also occur between countries. To the extent that there is greater offshoring of tasks in less interactive occupations from metro areas than from non-metro areas, this provides a related explanation in terms of the same mechanism for the increased concentration of employment in interactive occupations in metro areas relative to non-metro areas over time.

<sup>9</sup>Metro areas are defined in IPUMS based on Census Bureau Metropolitan Statistical Areas (MSAs).

individuals by their person weights to ensure the representativeness of the sample. Our main dataset is a panel from 1880-2000 that uses information on the share of employment within an occupation and sector in metro areas, for which the 1 percent IPUMS samples are representative.

We use the standardized 1950 occupation classification from IPUMS, which distinguishes eleven two-digit occupations (e.g. “Clerical and Kindred”) and 281 three-digit occupations (e.g. “Opticians and Lens Grinders and Polishers”). We also use the standardized 1950 sector classification from IPUMS, which distinguishes twelve two-digit sectors (e.g. “Finance, Insurance and Real Estate”) and 158 three-digit sectors (e.g. “Motor Vehicles and Motor Vehicle Equipment”).<sup>10</sup> Since we are concerned with employment structure, we omit workers who do not report an occupation or a sector (e.g. because they are unemployed or out of the labor force). We also exclude workers in agricultural occupations or sectors, because we compare task specialization in urban and rural areas over time, and agriculture is unsurprisingly overwhelmingly located in rural areas.<sup>11</sup> While our baseline sample uses time-varying boundaries of metro areas to ensure that these correspond to meaningful economic areas, we also report robustness checks using administrative cities whose boundaries are more stable over time.

Our second main data source is the Dictionary of Occupational Titles (U.S. Department of Labor 1991), which contains detailed descriptions of more than 12,000 occupations. Following Autor et al. (2003), previous research using DOTs typically uses the numerical scores that were constructed for each occupation by the Department of Labor (e.g. a Non-routine Interactive measure based on the Direction, Control and Planning of Activities (DCP) numerical score). In contrast, we use verbs from the detailed occupational descriptions in DOTs to directly measure the tasks performed by workers in each occupation. We use a list of over 3,000 English verbs from “Writing English,” a company that offers English language consulting.<sup>12</sup> This approach enables us to provide a detailed analysis of the multiple tasks undertaken by workers within each occupation without being restricted to a single numerical score for each occupation. We match the DOTs occupations to the three-digit occupations in our census data using the crosswalk developed by Autor, Levy and Murnane (2003). In our baseline specification, we use a time-invariant measure of tasks based on the occupational descriptions from the digital edition of the 1991 DOTs, which ensures that our results are not driven by changes in language use over time. In sensitivity checks, we also report results using digitized occupational descriptions from the first edition of the DOTs in 1939 (U.S. Department of Labor 1939). As an external validity check, we also compare our measures of occupational characteristics based on verbs from the DOTs to separate independent measures of occupational characteristics.

We complement these two main data sources with information from a variety of other sources. We use

---

<sup>10</sup>See IPUMS for the full concordance between two-digit and three-digit occupations and sectors. While both occupation and sector classifications are standardized by IPUMS, there are a small number of occupations and sectors that enter and exit the sample over time. All our results are robust to restricting attention to occupations and sectors that are present in all years.

<sup>11</sup>Our key findings, however, are robust to the inclusion of these agricultural workers. For further analysis of the relationship between urbanization and structural transformation away from agriculture, see Michaels et al. (2012).

<sup>12</sup>See <http://www.writingenglish.com/englishverbs.htm>.

the standard reference for word usage in English (Roget’s Thesaurus) to quantify the meanings of verbs from the occupational descriptions.<sup>13</sup> We use ArcGIS shapefiles from the National Historical Geographical Information System (NHGIS) to track the evolution of county boundaries over time.

## 4 Specialization Across Occupations and Sectors

We begin by providing motivating evidence of changes in specialization across occupations and sectors in metro areas relative to non-metro areas. To do so, we estimate the following regression for each year  $t$  separately using data across occupations  $o$  and sectors  $s$ :

$$\text{MetroShare}_{ost} = \mu_{ot} + \eta_{st} + \varepsilon_{ost}, \quad (20)$$

where  $\text{MetroShare}_{ost}$  is the share of employment in metro areas in occupation  $o$ , sector  $s$  and year  $t$ ; observations are weighted by person weights;  $\mu_{ot}$  are occupation-year fixed effects;  $\eta_{st}$  are sector-year fixed effects; and  $\varepsilon_{ost}$  is a stochastic error. We normalize the sector-year and occupation-year fixed effects so that they each sum to zero in each year, and hence they capture deviations from the overall mean in each year. We estimate the above regression using a share as the left-hand side variable so that the estimated coefficients have a natural interpretation as frequencies. We also find a very similar pattern of results in a robustness test in which we use a logistic transformation of the left-hand side variable:  $\text{MetroShare}_{ost}/(1 - \text{MetroShare}_{ost})$ .

The occupation-year fixed effects ( $\mu_{ot}$ ) capture the average probability of being in a metro area for workers in each occupation in each year, after controlling for differences across sectors in metro probabilities. Similarly, the sector-year fixed effects ( $\eta_{st}$ ) capture the average probability of being located in a metro area for workers in each sector in each year, after controlling for differences across occupations in metro probabilities. The sector and occupation fixed effects are separately identified because there is substantial overlap in occupations and sectors, such that each sector contains multiple occupations and each occupation is employed in several sectors.<sup>14</sup> We estimate this regression using both the aggregate (two-digit) and disaggregate (three-digit) definitions of occupations and sectors discussed above.

As reported in Table 1 for two-digit occupations and sectors, we find substantial changes in specialization across occupations and sectors in metro areas relative to non-metro areas over time. From Panel A, in 1880, “Clerical and Kindred” workers were the most likely to be located in metro areas. In contrast, by 2000, “Clerical and Kindred” workers were ranked only fourth, and “Professional and Technical” workers were the most likely to be located in metro areas. From 1880-2000, declines in ranks were observed for “Craftsmen” (from 2 to 6) and “Operatives” (from 3 to 7), while increases in ranks were observed for “Professional and Technical” workers (from 7 to 1) and “Managers, Officials and Proprietors” (from 6 to 3). As

<sup>13</sup>We use the online computer-searchable edition of Roget (1911): <http://machaut.uchicago.edu/rogets>.

<sup>14</sup>The average three-digit sector employs workers from 111 three-digit occupations, while the average three-digit occupation contains workers employed in 81 sectors.

apparent from the first and fourth columns of the table, these changes in ranks reflect substantial changes in the probabilities of workers in individual occupations being located in metro areas over time.

Since regression (20) includes sector-year fixed effects, these changes in the metro probabilities for each occupation are not driven by changes in sector composition, but rather reflect changes in the organization of economic activity within sectors. Nonetheless, we also observe substantial changes in sector structure in metro areas relative to non-metro areas over time. From Panel B, declines in ranks from 1880-2000 were observed for “Wholesale and Retail Trade” (from 2 to 6) and “Manufacturing” (from 4 to 10). In contrast, increases in ranks from 1880-2000 were observed for “Transportation, Communication and Other Utilities” (from 6 to 3) and “Business and Repair Services” (from 9 to 1).

In Figures A1 and A2 of the web appendix, we show the evolution of the occupation and sector coefficients across each of the twenty-year intervals in our data. While “Professional and Technical” workers display an increased propensity to locate in metro areas from 1880-1960, the probability that “Managers, Officials and Proprietors” are located in urban areas increases particularly sharply from 1940-2000. In contrast, the likelihood that “Craftsmen” are found in metro areas declines throughout our sample period, while the probability for “Clerical and Kindred” workers declines from 1900 onwards, and the probability for “Service” workers initially rises until 1920 and later declines until around 1960.

Such changes in specialization are not limited to the aggregate categories considered so far, but are also found using more disaggregated measures of occupations and sectors. In Table A1 of the web appendix, we report the results of estimating the regression (20) including three-digit-occupation-year and three-digit-sector-year fixed effects. Panels A and B report the twenty occupations within the largest increases and decreases respectively in the within-sector probability of being located in a metro area from 1880-2000. Both the top agglomerating occupations in Panel A and the top dispersing occupations in Panel B are diverse and span multiple sectors. For example, the top agglomerating occupations include “Editors and Reporters”, “Judges and Lawyers” and “Pattern and Model Makers,” while the top dispersing occupations contain “Office Machine Operators” and “Upholsterers.” In our empirical analysis below, we measure the multiple tasks undertaken by workers within each occupation, and provide evidence on the systematic characteristics shared by occupations that agglomerate versus disperse over time.

## **5 Measuring the Tasks Undertaken Within Occupations**

To measure the tasks undertaken by workers in each occupation, we use the detailed descriptions of more than 12,000 disaggregated occupations included in the DOTs. We use the verbs from each occupation’s description to measure the tasks performed by workers within that occupation, because verbs capture an action (bring, read, walk, run, learn), an occurrence (happen, become), or a state of being (be, exist, stand), and hence capture the task being performed. To focus on persistent characteristics of occupations and abstract

from changes in word use over time, our baseline analysis uses time-invariant occupational descriptions from the 1991 digital edition of the DOTs. While the tasks undertaken within each occupation can change over time, the relative task content of occupations is likely to be more stable. To provide evidence on the extent to which this is the case, we have also digitized the occupational descriptions from the first edition of the DOTs in 1939. Although the descriptions of occupations are less detailed in the historical DOTs, we find a similar pattern of results using both sets of occupational descriptions and provide evidence below on the correlation of the relative task content of occupations over time.

The first step of our procedure uses a list of over 3,000 English verbs from “Writing English,” a company that offers English language consulting. Using this list of verbs, we search each occupational description in the 1991 DOTs for occurrences of each verb in the first-person singular (e.g. (I) talk), third-person singular (e.g. (she) talks) or present participle (e.g. (he is) talking). To take an example from first-hand experience, the occupational description for an Economist is:

“ECONOMIST: *Plans, designs, and conducts research* to aid in interpretation of economic relationships and in solution of problems *arising* from production and distribution of goods and services: *Studies* economic and statistical data in area of specialization, such as *finance, labor, or agriculture*. *Devises* methods and procedures for *collecting* and *processing* data, *utilizing* knowledge of available sources of data and various econometric and *sampling* techniques. *Compiles* data *relating* to *research* area, such as employment, productivity, and *wages* and hours. *Reviews* and *analyzes* economic data in *order* to *prepare reports* detailing results of investigation, and to *stay* abreast of economic *changes ...*”

where the words detected by our procedure as capturing the tasks performed by an economist are italicized.<sup>15</sup> Note that sometimes the first-person singular, third-person singular or present participle forms of a verb have the same spelling as the corresponding adjectives and nouns (e.g. “prepare *reports*”). In this case, our procedure treats these adjectives and nouns as verbs. To the extent that the use of the same word as an adjective or noun is closely related to its use as a verb, both uses are likely to capture the tasks performed.

From this first step, we obtain the number of occurrences of each verb for each DOTs occupation. We next match the more than 12,000 DOTs occupations to IPUMS standardized 1950 occupations using the crosswalk developed by Autor, Levy and Murnane (2003). Finally, we calculate the frequency with which each verb  $v$  is used for each IPUMS occupation  $o$ :

$$\text{VerbFreq}_{vo} = \frac{\text{Appearances of verb } v \text{ matched to } o}{\text{Appearances of all verbs matched to } o},$$

where we focus on the frequency rather than the number of verb uses to capture the relative importance of

---

<sup>15</sup>As an indication of the wide coverage of our list of over 3,000 verbs, only 1,830 of these verbs appear in the 1991 DOTs occupational descriptions.

tasks for an occupation and to control for potential variation in the length of the occupational descriptions matched to each IPUMS occupation.

We provide evidence on changes in task specialization in metro areas relative to non-metro areas over time by estimating the following regression for each verb  $v$  and year  $t$  separately using data across occupations  $o$  and sectors  $s$ :

$$\text{MetroShare}_{ost} = \alpha_{vt} \text{VerbFreq}_{vo} + \eta_{vst} + \varepsilon_{ost}, \quad (21)$$

where  $\text{MetroShare}_{ost}$  is again the share of employment in metro areas in occupation  $o$ , sector  $s$  and year  $t$ ;  $\text{VerbFreq}_{vo}$  is defined above for verb  $v$  and occupation  $o$ ;  $\eta_{vst}$  are verb-sector-year fixed effects; and  $\varepsilon_{ost}$  is a stochastic error.

The coefficient of interest  $\alpha_{vt}$  captures a conditional correlation: the correlation between occupations' shares of employment in metro areas and their frequency of use of verb  $v$ . The verb-sector-year fixed effects ( $\eta_{vst}$ ) control for differences across sectors in the frequency of verb use and for differences across sectors and over time in the concentration of employment in metro areas. Since  $\text{VerbFreq}_{vo}$  is time invariant, a rise in  $\alpha_{vt}$  over time implies that employment in occupations using that verb is increasingly concentrating in metro areas within sectors over time.

In Panels A and B of Table 2, we report for each year the ten verbs with the highest and lowest standardized coefficient  $\alpha_{vt}$  (the estimated coefficient multiplied by the standard deviation of  $\text{VerbFreq}_{vo}$ ).<sup>16</sup> As apparent from Panel A, we find substantial changes in the tasks most concentrated in metro areas within sectors over time. In 1880, the verbs with the highest metro employment shares typically involve physical tasks such as “Braid,” “Sew,” “Stretch” and “Thread.” By 1920, the top ten verbs include an increased number of clerical tasks, such as “Bill,” “File,” “Notice,” and “Record.” By 1980 and 2000, the leading metro verbs include a proliferation of interactive tasks, such as “Analyze,” “Advise,” “Confer” and “Report.” As shown in Panel B, we also find some changes in the tasks least concentrated in metro areas, although here the pattern is less clear cut (e.g. “Tread” appears from 1880-1960 and “Turn” appears from 1960-2000).

## 6 Quantifying Task Specialization

The approach developed in the previous section allows us to provide a detailed characterization of the tasks performed in urban and rural areas using the 3,000 verbs and 12,000 occupational descriptions. In this section, we use the meanings of these verbs to quantify the way in which the task composition of employment has evolved over time. In particular, we use an online computer-searchable version of Roget's Thesaurus (1911), which has been the standard reference for English language use for more than a century and explicitly classifies words according to their underlying concepts and meanings. Roget's classification

<sup>16</sup>We find a similar pattern of results just using the estimated coefficients instead of the estimated coefficients times the standard deviation of  $\text{VerbFreq}_{vo}$ .

was inspired by natural history, with its hierarchy of Phyla, Classes, Orders and Families. Therefore words are grouped according to progressively more disaggregated classifications that capture ever more subtle variations in meaning. A key advantage of this classification is that it explicitly takes into account that words can have different meanings depending on context by including extensive cross-references to link related groups of words.<sup>17</sup>

Roget’s Thesaurus is organized into six “Classes” that are further disaggregated into the progressively finer subdivisions of “Divisions,” “Sections” and “Categories.” The first three classes cover the external world: Class I (Abstract Relations) deals with ideas such as number, order and time; Class II (Space) is concerned with movement, shapes and sizes; and Class III (Matter) covers the physical world and humankind’s perception of it by means of the five senses. The last three classes relate to the internal world of human beings: the human mind (Class IV, Intellect), the human will (Class V, Volition) and the human heart and soul (Class VI, Emotion, Religion and Morality).

To characterize the meaning of each verb  $v$ , we use the frequency with which it appears in each subdivision  $k$  of Roget’s Thesaurus:

$$\text{ThesFreq}_{vk} = \frac{\text{Appearances of verb } v \text{ in subdivision } k \text{ of thesaurus}}{\text{Total appearances of verb } v \text{ in thesaurus}}, \quad (22)$$

where our use of a frequency takes into account that each verb can have multiple meanings and provides a measure of the relative importance of each meaning. In counting the appearances of verbs we make use of the thesaurus’s structure, in which words with similar meanings appear under each thesaurus Category in a list separated by commas or semi-colons. Based on this structure, we count appearances of a verb that are followed by a comma or semi-colon, which enables us to abstract from appearances of a word in idioms that do not reflect its common usage.<sup>18</sup>

Combining the frequency with which a verb appears in each occupation’s description ( $\text{VerbFreq}_{vo}$  in the previous section) and the frequency with which the verb appears in each subdivision of the thesaurus ( $\text{ThesFreq}_{vk}$ ), we construct a quantitative measure of the extent to which the tasks performed in an occupation involve the concepts from each thesaurus subdivision:

$$\text{TaskContent}_{ko} = \sum_{v \in V} \text{VerbFreq}_{vo} \times \text{ThesFreq}_{vk}.$$

We use this measure to examine changes in task specialization in metro areas relative to non-metro areas over time by estimating an analogous regression for each thesaurus subdivision  $k$  and year  $t$  as for each verb and year in the previous section:

$$\text{MetroShare}_{ost} = \beta_{kt} \text{TaskContent}_{ko} + \eta_{kst} + \varepsilon_{ost}, \quad (23)$$

<sup>17</sup>For further discussion of the genesis of Roget’s Thesaurus, see for example Hüllen (2003).

<sup>18</sup>For example, the verb “Consult” appears in six thesaurus Categories. The entry followed by a comma is 695 Advice, which captures the word’s meaning. Entries not followed by a comma correspond to idiomatic uses not closely related to the word’s meaning: 133 Lateness (“consult one’s pillow”); 463 Experiment (“consult the barometer”); 707 Aid (“consult the wishes of”); 943 Selfishness (“consult one’s own pleasure”); 968 Lawyer (“juris consult [Latin]”).

where  $\text{MetroShare}_{ost}$  is the share of employment in metro areas in occupation  $o$ , sector  $s$  and year  $t$ ;  $\text{TaskContent}_{ko}$  is defined above for thesaurus subdivision  $k$  and occupation  $o$ ;  $\eta_{kst}$  are thesaurus-subdivision-sector-year fixed effects; and  $\varepsilon_{ost}$  is a stochastic error.

The coefficient of interest  $\beta_{kt}$  again captures a conditional correlation: the correlation between occupations' shares of employment in metro areas and their frequency of use of verbs in thesaurus subdivision  $k$ . The thesaurus-subdivision-sector-year fixed effects ( $\eta_{kst}$ ) control for differences across sectors in the frequency of use of thesaurus subdivisions and differences across sectors and over time in the concentration of employment in metro areas. Since  $\text{TaskContent}_{ko}$  is time invariant, a rise in  $\beta_{kt}$  over time implies that employment in occupations using that subdivision of the thesaurus is increasingly concentrating in metro areas within sectors over time.

In Table 3, we report the estimation results for the thirty-eight Sections of the thesaurus. We calculate the standardized coefficient for each Section of the thesaurus (the estimated coefficient  $\beta_{kt}$  multiplied by the variable's standard deviation) and report the ranking of these standardized coefficients in 1880 and 2000 as well the difference in rankings between these two years (1880 minus 2000).<sup>19</sup> Since the thesaurus Section with the highest standardized coefficient is assigned a rank of one, positive differences in rankings correspond to thesaurus categories that are becoming more concentrated in metro areas within sectors over time. The table highlights the top-five thesaurus Sections in 1880 in bold-italics and the top-five thesaurus Sections in 2000 in bold.

The results in Table 3 reveal a sharp change the relative ranking of thesaurus Sections involving the external world (Classes I-III) and those involving the internal world of human beings (Classes IV-VI). In 1880, the top-five thesaurus Sections most concentrated in metro areas were: Quantity (Class I), Time (Class I), Matter in General (Class III), Dimensions (Class II), and Inorganic Matter (Class III). In contrast, in 2000, the top-five thesaurus Sections were: Results of Reasoning (Class IV), Means of Communicating Ideas (Class V), Moral Affections (Class VI), Voluntary Action (Class IV) and Precursory Conditions and Operations (Class IV). The correlation between the rankings of the thesaurus sections in 1880 and 2000 is negative and statistically significant (-0.43).

This approach of using the thesaurus Sections to capture the meanings of verbs enables us to characterize non-parametrically the way in which in the task composition of employment in metro and non-metro areas has changed over time. We find that positive changes in ranks in Table 3 are typically concentrated in thesaurus Classes IV and V, which includes Class IV, Division 1 (Formation of Ideas), Class IV, Division 2 (Communication of Ideas) and Class V, Division 2 (Intersocial Volition). Therefore the positive changes in ranks reflect an increased concentration of tasks involving thought, communication and intersocial activity in metro areas.

---

<sup>19</sup>Again we find a similar pattern of results using just the estimated coefficient instead of the estimated coefficient times the standard deviation of  $\text{TaskContent}_{ko}$ .

We define this combination of thought, communication and intersocial activity as “interactiveness.” An example of such interaction is a business meeting, presentation, seminar or conference. Each of these examples involves the creation and transformation of ideas (thought), the expression of these potentially complex ideas (communication), and the understanding of these potentially difficult ideas by others (intersocial activity). We do not exclude some ideas being generated in isolation or some communication occurring without intersocial activity (e.g. in the form of written media such as memos and publications). But much knowledge is tacit and hard to codify; the understanding of complex ideas often requires considerable bandwidth and two-way dialogue; and face-to-face interactions can be important in conveying subtleties of emphasis and meaning. While innovations in communication technology such as the telephone and the internet can be important in reducing the costs of communicating at a distance, it remains the case that considerable amounts of time and other resources are devoted by businesses and other organizations to bringing individuals together to interact with one another.

We construct our baseline measure of the overall interactiveness of an occupation based on the frequency with which verbs appear in that occupation’s description and the frequency with which those verbs appear in Classes IV and V of the thesaurus (which includes the formation of ideas, the communication of ideas and intersocial volition as discussed above):

$$\text{Interactive}_o = \sum_{v \in V} \text{FreqVerb}_{vo} \times \text{FreqInteractive}_v, \quad (24)$$

where  $\text{FreqVerb}_{vo}$  is the frequency with which verb  $v$  is used for occupation  $o$  from above;  $\text{FreqInteractive}_v$  is the frequency with which verb  $v$  appears in thesaurus Classes IV and V (computed as in (22)). We also report results below for the separate components of interactiveness – thought (Class IV, Division 1), communication (Class IV, Division 2) and intersocial activity (Class V, Division 2) – and show that all three components play an important role. Finally, we compare our measure of the interactiveness of occupations based on the meanings of verbs to separate independent measures, as discussed further in subsection 7.4 below.

In Panels A and B of Table 4, we report the top ten and bottom ten interactive occupations based on our measure. While any single quantitative measure of interactiveness is unlikely to fully capture the meaning of this concept, the occupations identified by our procedure as having high and low levels of interactiveness appear intuitive. Interactive tasks are arguably more central to the set of tasks performed by “Buyers and Department Heads”, “Clergymen” and “Pharmacists” than they are to the set of tasks performed by “Blasters and Powdermen”, “Roofers and Slaters” and “Welders and Flame Cutters.” Note that each occupation is likely to perform some interactive tasks: for example, “Roofers and Slaters” can negotiate with clients and liaise with architects. A key advantage of our measure is that it captures the relative importance of interactive tasks compared to all other tasks performed by workers within an occupation, because it equals the frequency with which verbs are used for an occupation multiplied by the frequency with which these

verbs appear in interactive subdivisions of the thesaurus. For example, while “Roofers and Slaters” can negotiate with clients and liaise with architects (Classes IV and V of the thesaurus), building and repairing roofs (Classes I-III of the thesaurus) is arguably more central to the set of tasks undertaken by workers within this occupation.

In Figure 1, we measure the interactiveness of metro areas, non-metro areas and the economy as a whole using the employment-weighted average of interactiveness for each occupation. In this measure, interactiveness only differs between metro and non-metro areas to the extent that they have different distributions of employment across occupations:

$$\text{Interactive}_{jt} = \sum_{o=1}^O \frac{E_{ojt}}{E_{jt}} \text{Interactive}_o, \quad j \in \{M, N\}, \quad (25)$$

where  $j$  indexes a type of location and we again denote metro areas by  $M$  and non-metro areas by  $N$ ;  $E_{ojt}$  corresponds to employment in occupation  $o$  in location type  $j \in \{M, N\}$  in year  $t$ .

As shown in the figure, metro and non-metro areas have similar levels of interactiveness in 1880. If anything, non-metro areas have higher interactiveness than metro areas. From the early decades of the twentieth century onwards, interactiveness increases in both sets of locations. This increase, however, is larger in metro areas than in non-metro areas, so that by the end of the sample period metro areas are substantially more interactive than non-metro areas.

## 7 Robustness

Having presented our baseline evidence on task specialization in metro and non-metro areas over time, we now document the robustness of these findings across a large number of different samples and specifications.

### 7.1 Metro Areas and Administrative Cities

Our analysis has so far used variation between metro and non-metro areas. To provide further evidence of a relative increase in the interactiveness of employment in densely-populated locations, we now present evidence using a different source of variation across metro areas of differing population densities.

In Figure 2, we display mean interactiveness for each metro area (as calculated using (25)) against log population density, as well as the fitted values and confidence intervals from locally-weighted linear least squares regressions. Panels A-C show results for 1880, 1940 and 2000 respectively. To ensure that metro areas correspond to meaningful economic units in each year, we use time-varying definitions of metro areas from IPUMS. Therefore both the boundaries of each metro area and the number of metro areas changes over time. In 1880, we find little relationship between interactiveness and log population density across metro areas, which is reflected in a negative but statistically insignificant OLS coefficient (standard error)

of -0.0002 (0.0013). In 1940, the relationship between interactivensness and log population density across metro areas remains relatively flat. In contrast, in 2000, we find a strong positive and statistically significant relationship between interactivensness and log population density, which is reflected in an OLS coefficient (standard error) of 0.0018 (0.0002). We find a similar pattern of results for 2000 if we restrict the sample to 1880 or 1940 metro areas, as shown in Panel D for 1940 metro areas. The magnitude of the difference in interactivensness between metro and non-metro areas in 2000 in Figure 1 is comparable to difference in interactivensness across metro areas of different population densities in 2000 in Figure 2.

Metro areas with relatively high levels of interactivensness conditional on population density in 2000 include Boston and New York, while those with low levels of interactivensness conditional on population density in 2000 include Anniston and Mansfield. These differences in interactivensness across metro area are persistent over time: the correlation coefficient between interactivensness in 1880 and 2000 is 0.52. Metro area interactivensness is also predictive of subsequent population growth: the correlation coefficient between 1880 interactivensness and 1880-2000 population growth is 0.33.<sup>20</sup>

As noted above, our baseline specification uses time-varying definitions of metro areas to ensure that they correspond to meaningful economic units in each year. One potential concern is that the change in the relative interactivensness of metro areas could be driven by a change in the geographical boundaries of metro areas as they have expanded to include surrounding suburbs. To address this concern, we replicated our analysis using an alternative definition of urban areas as administrative cities, which have much more stable geographical boundaries over time. Again we find an increase in the relative interactivensness of urban areas over time, whether we compare administrative cities to all other locations (Figure A4 in the web appendix) or only to non-metro areas (Figure A5 in the web appendix). We also find an increasingly positive relationship over time between interactivensness and population density across administrative cities of different sizes. Therefore our findings of an increase in the relative interactivensness of employment in more-densely-populated locations reflect a change in the organization of economic activity within existing geographical boundaries.

## 7.2 Other Interactivensness Measures

Our approach of using verbs from the occupational descriptions enables us to measure the multiple tasks undertaken by workers within each occupation and to construct a measure of the interactivensness of occupations based on the meanings of these verbs. In this section, we compare this measure to independently-constructed interactivensness measures.

In Table A3 of the web appendix, we report the correlation across occupations between our interactivensness measure and seventeen subcategories of the Occupational Information Network (O\*NET) work activity

---

<sup>20</sup>We find similar patterns comparing values in 1940 and 2000.

“Interacting with Others.” These measures were constructed by US Department of Labor/Employment and Training Administration (USDOL/ETA) based on questionnaires about detailed work activities issued to a random sample of businesses and workers. Panel A reports unweighted correlations, while Panel B reports correlations weighted by employment. The measures cover a wide range of forms of interaction, including “Assisting and caring for others” and “Resolving conflict and negotiating with others.”

We find that all of the correlations with our measure of interactiveness are positive and statistically significant. The five categories with the highest unweighted correlations correspond closely to activities involving thought, communication and intersocial activity: “Interpreting the meaning of information for others,” “Provide consultation and advice to others,” “Resolving conflict and negotiating with others,” “Establishing and maintaining interpersonal relationships” and “Performing administrative activities.” These positive correlations provide external validation that our measure of interactiveness based on the meaning of verbs is capturing work activities thought to be interactive by businesses and their employees.

Our approach of using the verbs from the occupational descriptions has a number of advantages relative to these alternative measures. It permits an analysis of the multiple tasks undertaken by workers within each occupation (e.g. “Analyze,” “Advise,” “Confer” and “Report”), allows the data to reveal the nature of these tasks through the meanings of the verbs (without imposing a prior measure of a single occupational characteristic on the data), and can be undertaken using both historical and contemporary occupational definitions (rather than being restricted to contemporary measures).

### **7.3 1939 DOTs**

Our baseline specification measures the task content of employment using time-invariant occupational descriptions from the 1991 DOTs. While this approach ensures that our findings are not driven by changes in language use over time, it assumes that the relative task content of occupations is persistent over time. One concern is that the interactiveness of occupations could have changed over time and these changes in interactiveness could be correlated with occupations’ shares of employment in metro areas.

To address this concern, we replicated our analysis using the first edition of the DOTs from 1939. We digitized the more than 12,000 occupational descriptions in the 1939 DOTs and implemented our procedure of searching for verbs in each occupational description. The occupational descriptions are less detailed in the 1939 DOTs, which implies that the resulting measures of the task content of employment are likely to be less precise than those using the 1991 DOTs. Nonetheless, as reported in Table A2 of the web appendix, we find similar changes in task specialization in this robustness test. The verbs most correlated with metro employment shares in 1880 include physical tasks such as “Slot,” “Thread,” “Straighten” and “Stitch.” In contrast, the verbs most correlated with metro employment shares in 2000 include interactive tasks such as “Advise,” “Present,” “Question” and “Report.”

Using the verbs from the 1939 occupational descriptions and the frequency with which these verbs appear in Classes IV and V of the thesaurus, we again find an increase in the interactiveness of employment over time that is more rapid in metro areas than in non-metro areas, as shown in Figure A3 in the web appendix. This similarity of the results using both the 1939 and 1991 occupational descriptions suggests that our findings are unlikely to be driven by changes in the relative interactiveness of occupations over time. Indeed, although the layout of the occupational descriptions implies that our measure of interactiveness using the 1939 DOTs is less precise than our baseline measure using the 1991 DOTs (which by itself would induce an imperfect correlation), we find that they are positively and statistically significantly correlated. As reported in Table A4 of the web appendix, the unweighted correlation coefficient between the 1939 and 1991 measures across the sample of occupations in 2000 is 0.622.

## 7.4 Other Occupational Characteristics

In Table A4 of the web appendix, we also examine the correlation between our interactiveness measure and other occupation characteristics. We consider the five numerical scores for each occupation from the 1991 DOTs that were constructed by the Department of Labor and used by Autor, Levy and Murnane (2003). We find that our interactiveness measure has a negative or statistically insignificant correlation with Routine Cognitive (STS), Routine Manual (FINGER) and Nonroutine Manual (EHF). We find positive and statistically significant correlations for Non-routine Interactive (DCP) and Non-routine Analytic (GED-MATH). While both of these measures are related to the concepts of thought, communication and intersocial activity captured by our interactiveness measure, the correlations are around 0.5. Therefore our interactiveness measure captures distinctive information about the tasks performed by workers within occupations. While DCP is orientated towards top-down interactions between workers (e.g. between a manager and her subordinates), our measure captures all interactions between workers (e.g. between members of a product design team). While GED-MATH is orientated towards thought, our measure of interactiveness also captures communication and intersocial activity.

In Figure A6 in the web appendix, we show the relationship across occupations between our interactiveness measure and a measure of education based on the share of workers with a college degree. We display this relationship for each year in our sample after 1940 (the first year for which education information is available in our data). Although there is a positive correlation between our interactiveness measure and college education, high-interaction occupations can have both low and high levels of education. Repeating the analysis in Figure 1 and graphing mean interactiveness in metro and non-metro areas for low-education occupations only (those with a college-educated share of less than 0.2 in either 1940 or 2000 in Figure A6), we again find an increased interactiveness of employment in metro areas, as shown in Figure A7 of the web appendix. Therefore the concentration of interactive occupations in metro areas not simply capturing a

concentration of high-education occupations in metro areas.

Nonetheless the changes in the occupational and educational composition of employment are related in interesting ways. As also shown in Figure A6, the positive correlation between interactiveness and education strengthens over time. Therefore the expansion in the share of college-educated workers is non-neutral across occupations and is more concentrated in high-interactiveness occupations. The growth in the college-educated share is also concentrated in metro areas and we find an important interaction between metro areas and interactive occupations. Between 1940 and 2000, the difference in the college-educated share between metro and non-metro areas increases from 1 to 11 percentage points, with 74 percent of this increase concentrated in occupations with above-median levels of interactiveness. Therefore changes in the interactiveness of employment in metro and non-metro areas play an important role in understanding human capital differences between cities and the growth in levels of human capital over time. The increase in interactiveness, however, is not simply capturing an increase in human capital, because as shown above we find an increased concentration of interactive occupations in metro areas even in occupations that continue to have low levels of college education at the end of our sample period.

## 8 Explaining Changes in Interactiveness

Having demonstrated the robustness of the increase in the relative interactiveness of metro areas across a number of different samples and specifications, we now provide further evidence on explanations for the observed change in interactiveness. First, we decompose the overall change in interactiveness into the contributions of individual occupations and sectors, which enables us to explore explanations that emphasize particular occupations and sectors. Second, we report regression specifications using variation in interactiveness between sectors, within sectors, and within sectors and occupations over time. Using these regressions, we explore the importance of the constituent components of interactiveness (thought, communication and intersocial) and present evidence on a number of potential explanations.

### 8.1 Decomposing Interactiveness

We begin by decomposing the change in the overall interactiveness of metro and non-metro areas into the contributions of each two-digit occupation and sector. Overall interactiveness for metro and non-metro areas is the employment-weighted average of interactiveness for each two-digit-sector-occupation cell:

$$I_{jt} = \sum_{z \in \Omega} \sum_{o \in \Omega_z} \frac{E_{ojt}}{E_{jt}} I_o, \quad j \in \{M, N\}, \quad (26)$$

where  $z$  indexes two-digit-sector-occupation cells;  $o$  indexes disaggregated three-digit occupations within these cells; and  $t$  indexes time;  $\Omega$  is the set of two-digit-sector-occupation cells;  $\Omega_z$  is the set of three-digit

occupations within each cell  $z$ ; the interactivensness of each three-digit occupation is measured using (24) based on the time-invariant occupational descriptions from the 1991 DOTs.

Taking differences between times  $T$  and  $t > T$ , the change in the overall interactivensness of metro and non-metro areas can be decomposed as follows:

$$\Delta I_{jt} = \sum_{z \in \Omega} \sum_{o \in \Omega_z} \left[ \Delta \left( \frac{E_{ojt}}{E_{jt}} \right) \right] I_o, \quad j \in \{M, N\}, \quad (27)$$

where  $\Delta I_{jt} = I_{jt} - I_{jT}$ ;  $\Delta (E_{ojt}/E_{jt})$  is the change in the employment share of occupation  $o$  in location  $j \in \{M, N\}$ ; and we have used the fact that occupation interactivensness is constant over time. Taking differences again between metro and non-metro areas, we obtain an analogous decomposition of the change in the relative interactivensness of metro and non-metro areas:

$$\Delta I_{Mt} - \Delta I_{Nt} = \sum_{z \in \Omega} \sum_{o \in \Omega_z} \left[ \Delta \frac{E_{oMt}}{E_{Mt}} - \Delta \frac{E_{oNt}}{E_{Nt}} \right] I_o, \quad (28)$$

where the right-hand sides of the decompositions (27) and (28) are summations over the contributions from each two-digit-sector-occupation-cell. These contributions correspond to a matrix with two-digit sectors for rows and two-digit occupations for columns, where the right-hand side is a summation across both rows and columns. Metro areas display a larger increase in interactivensness than non-metro areas to the extent that they experience a greater reallocation of employment shares towards high-interactivensness occupations.

Figure 3 summarizes the results from the decompositions of the change in the relative interactivensness of metro and non-metro areas from (28). Panels A and B show the contributions for each two-digit occupation (summing across sectors in the rows of the matrix of contributions) for each twenty-year interval in our sample, while Panels C and D show the corresponding contributions for each two-digit sector (summing across occupations in the columns of the matrix of contributions).<sup>21</sup> Figures A8 and A9 in the web appendix report analogous results from the decompositions of the change in interactivensness for metro and non-metro areas separately from (27).

Panels A and B of Figure 3 show that the sharp increase in the relative interactivensness of metro areas from 1880-1920 is largely driven by positive contributions from Clerks (and to a lesser extent Professionals), with Operatives, Sales Workers and Managers all making negative contributions. From 1920-1960, Professionals (and to a lesser but growing extent Managers) make the largest positive contributions, while Craftsmen and Service Workers make negative contributions. From 1960-2000, Professionals and Managers have the largest positive contributions, while Clerks have the largest negative contribution.

Panels C and D of Figure 3 show that Professional and Business services are the two sectors that make the largest contributions to the increase in the relative interactivensness of metro areas over the sample as

<sup>21</sup>Since the change in overall interactivensness is the sum across all elements in the matrix, adding the sums for occupations and the sums for sectors would result in double-counting (since each element would be counted twice).

a whole. Professional Services are more important earlier in the sample period, while Business Services become more important later on. The sector that makes the largest negative contribution over the sample period as a whole is Wholesale and Retail trade, with the absolute magnitude of its contribution diminishing over time. While the contribution from Manufacturing is initially positive (up to 1920), it becomes negative from 1940 onwards.

Taking these decomposition results together, the increase in the relative interactiveness of metro areas is not driven by any one occupation or sector. Our results are not solely explained by Managers (whose contribution only becomes positive towards the end of our sample period). Clerks and Professionals make notable positive contributions towards the beginning and end of our sample period respectively. Our results are also not simply driven by a decline of Manufacturing in urban areas (indeed Manufacturing was expanding in the early decades of our sample when some of the largest changes in interactiveness were observed). Similarly, our findings are not simply attributable to an expansion of Services in urban areas (indeed Services was a smaller share of employment in the early decades of our sample when some of the largest changes in interactiveness were observed). Instead we find evidence of a pervasive reallocation of employment towards more interactive occupations within sectors.

## 8.2 Variation Within and Between Sectors

To further explore the determinants of the increase in the relative interactiveness of metro areas, we begin by examining between-sector variation. Sector interactiveness is measured as the employment-weighted mean of the interactiveness of each occupation:

$$\text{Interactive}_{st} = \sum_o \frac{E_{ost}}{E_{st}} \text{Interactive}_o,$$

We run a regression across sectors of the share of a sector's employment in metro areas ( $\text{MetroShare}_{st}$ ) on its interactiveness ( $\text{Interactive}_{st}$ ) for each year separately:

$$\text{MetroShare}_{st} = \alpha_t \text{Interactive}_{st} + \varepsilon_{st}, \quad (29)$$

where  $\varepsilon_{st}$  is a stochastic error;  $\alpha_t$  captures the correlation between sectors' shares of employment in metro areas and their interactiveness in each year. We estimate the above regression and the remaining regressions in this section using a share as the left-hand side variable so that the estimated coefficients have a natural interpretation as frequencies. But we find a very similar pattern of results in a robustness test in which we use a logistic transformation of the left-hand side variable:  $\text{MetroShare}_{st}/(1 - \text{MetroShare}_{st})$ .

Panel A of Table 5 reports the results, where each cell in the table corresponds to a separate regression. The first four rows are based on metro employment shares (as in the regression above and shown in parentheses in the table), while the final row is based on metro wage bill shares (as discussed further below

and shown in parentheses in the table). From the first row of the table, there is a negative but statistically insignificant correlation between a sector's metro employment share and its interactiveness in 1880. From 1900 onwards, there is an increase in the correlation between a sector's metro employment share and its interactiveness, which is particularly sharp from 1900-1940, and becomes positive and statistically significant at conventional critical values in 1960. Therefore more interactive sectors become increasingly concentrated in metro areas over time.

In the second to fourth rows of Panel A of Table 5, we also break out overall interactiveness into thought (Class IV, Division 1), communication (Class IV, Division 2) and intersocial (Class V, Division 2). As shown in the table, we find that the sectors increasingly concentrating in metro areas over time involve each of these components of interactiveness. Therefore the between-sector rise in the interactiveness of employment in metro areas is driven by the combination of thought, communication and intersocial activity.

These changes in patterns of specialization in metro versus non-metro areas are explained in the model by changes in the relative demand for occupations as a result of either differential changes in the relative productivity of occupations or improvements in communication technology. An increase in the relative demand for an occupation raises both its employment and its wage (and hence raises its wagebill). In contrast, an increase in the relative supply of an occupation raises its employment but reduces its wage (and hence reduces its wagebill if the demand for occupations is inelastic). To assess the relative importance of these two explanations, the final row of Panel A of Table 5 reports the results of regressions in which we use the share of the sector's wagebill in metro areas (rather than its share of employment in metro areas) as the left-hand side variable. Although the wage data are available for a much shorter time period than the employment data, we find a similar pattern of results using this alternative left-hand side variable, which is consistent with relative demand moving relative wagebills and employment in the same direction.

Having established these relationships between sectors, we next examine within-sector variation. We run a regression across sectors and occupations of the share of a sector-occupation's employment in metro areas ( $\text{MetroShare}_{ost}$ ) on occupation interactiveness ( $\text{Interactive}_o$ ) for each year separately:

$$\text{Metro}_{ost} = \alpha_t \text{Interactive}_o + \eta_{st} + \varepsilon_{ost}, \quad (30)$$

where  $\eta_{st}$  are sector-year fixed effects and  $\varepsilon_{ost}$  is a stochastic error. The sector-year fixed effects ( $\eta_{st}$ ) control for changes in sector composition over time, so that the coefficient  $\alpha_t$  is identified solely from variation within sectors. The coefficient  $\alpha_t$  captures the within-sector correlation between the share of employment in metro areas and the interactiveness of occupations.

Panel B of Table 5 reports the results, where each cell in the table again corresponds to a separate regression. The first four rows are again based on metro employment shares, while the final row is based on metro wage bill shares. As shown in the first row, and in line with our previous results, the correlation between metro employment shares and interactiveness is negative in 1880. Over time, this correlation becomes more

positive and becomes statistically significant by 1960. Therefore, within sectors, more interactive occupations become increasingly concentrated in metro areas over time. This finding of the same pattern of reallocation across occupations both between and within sectors is consistent with a wide-ranging secular process favoring specialization in interactive occupations in metro areas.

In the second to fourth rows of Panel B of Table 5, we also break out overall interactiveness into thought (Class IV, Division 1), communication (Class IV, Division 2) and intersocial (Class V, Division 2). Again we find that the increased interactiveness of employment in metro areas is driven by the combination of thought, communication and intersocial activity. In the final row of Panel B of Table 5, we report the results of regressions in which we use the share of a sector-occupation's wagebill in metro areas (rather than its share of employment in metro areas) as the left-hand side variable. For the shorter period over which we have the wage data, we again find a similar pattern of results using this alternative left-hand side variable, which is consistent with relative demand moving relative wagebills and employment in the same direction.

Finally, to use variation within sectors and occupations, we pool our sector-occupation data over time and estimate a panel data regression that facilitates the inclusion of sector, occupation and year fixed effects. We regress the share of a sector-occupation's employment in metro areas on these fixed effects and interaction terms between time dummies and our measure of occupation interactiveness:

$$\text{MetroShare}_{ost} = \alpha_t [\text{Interactive}_o \times \text{Year}_t] + \mu_o + \eta_s + \delta_t + \varepsilon_{ost}, \quad (31)$$

where  $\varepsilon_{ost}$  is a stochastic error; we choose 1880 as the excluded year from the interaction terms. The occupation fixed effects ( $\mu_o$ ) control for time-invariant differences between metro and non-metro areas in the share of an occupation in employment and capture the main effect of occupation interactiveness. The sector fixed effects ( $\eta_s$ ) control for time-invariant differences between metro and non-metro areas in the share of a sector in employment. The year fixed effects ( $\delta_t$ ) control for changes in the shares of metro areas in employment across all occupations and sectors. The coefficients  $\alpha_t$  capture the change in the correlation between metro employment shares and interactiveness relative to 1880.

Table 6 reports the estimation results. Column (1) confirms our findings above of an increasing correlation between metro employment shares and occupation interactiveness over time, which becomes positive and statistically significant by 1960. As shown in Column (2), this increasing correlation between metro employment shares and occupation interactiveness is robust to replacing the sector and year fixed effects with sector-year fixed effects to control for changes in sector composition over time.

While the model's explanation for the increase in the relative interactiveness of metro areas emphasizes changes in the relative demand for occupations as a result of either differential changes in the relative productivity of occupations or improvements in communication technology, Columns (3)-(4) consider an alternative explanation based on changes in female labor force participation. Over our long historical time period, female labor force participation increased substantially, which implies that more and more couples

face a colocation problem where both partners are looking for work in a common location (e.g. Costa and Kahn 2000). Since solving such a colocation problem is likely to be easier in more densely-populated locations, one concern is that the movement of such “power couples” into densely-populated locations could be driving the increase in the relative concentration of employment in interactive occupations in metro areas. Although it is not necessarily the case that power couples work in interactive occupations, Columns (3) and (4) provide evidence against this concern by estimating the specification in Column (2) separately for single and married people. Comparing the two columns, we find a similar pattern of results irrespective of marital status, which suggests that our findings are not being driven by the location decisions of power couples.

In Columns (5)-(7), we provide further evidence against explanations based on individual sectors and occupations. In Column (5), we include only workers in the manufacturing sector and demonstrate a similar pattern of results, which corroborates that our findings are not simply being driven by the rise of the services sector in urban areas. In Column (6), we include only workers in the services sector, which confirms that our findings are not simply being driven by a decline in manufacturing in urban areas. In Column (7), we exclude 22 three-digit occupations that are likely to be concentrated in headquarters (e.g. “Buyers and Department Heads,” “Clerical and Kindred Workers” and “Managers, Officials and Proprietors.”) After excluding these occupations, we continue to find an increasing interactiveness of employment in metro areas over time, suggesting that our results are not being driven by headquarters alone.

In Columns (8)-(9), we examine differences in human capital across cities. Glaeser and Resseger (2009) find that the positive average relationship between productivity and metro area population is driven by a strong positive relationship for more-skilled metro areas, whereas this relationship is almost non-existent for less-skilled metro areas. Using Glaeser and Resseger (2009)’s classification of metro areas by skill, Columns (8) and (9) re-estimate the specification in Column (2) excluding more and less-skilled metro areas respectively.<sup>22</sup> In both samples, we find a positive and statistically significant increase in the relative concentration of employment in interactive occupations in metro areas over time. The size of this increase is larger in the sample excluding less-skilled metro areas. But even in the sample excluding more-skilled metro areas we find a reallocation of employment towards interactive occupations in metro areas.

## 9 Conclusions

We introduce a new methodology for measuring the tasks undertaken within occupations. We use this methodology to provide the first evidence on task specialization in urban and rural areas in the United States over a long historical time period. We measure tasks using the verbs from occupational descriptions and interpret the meanings of these verbs using Roget’s Thesaurus. In 1880, the tasks most concentrated

---

<sup>22</sup>In Glaeser and Resseger (2009)’s classification, more-skilled Metropolitan Statistical Areas (MSAs) have a share of adults with college degrees of greater than 25.025 percent in 2006. The year 1960 is omitted in Columns (8) and (9) because the IPUMS 1960 data do not contain the identifiers for individual MSAs.

in metro areas were “Braid,” “Sew,” “Stretch” and “Thread.” In contrast, in 2000, those most concentrated in metro areas were “Analyze,” “Advise,” “Confer” and “Report.” We find a secular reallocation over time towards more interactive tasks, defined as those involving thought, communication and intersocial activity. At the beginning of our sample period, metro and non-metro areas had similar levels of interactiveness. From the early decades of the twentieth century onwards, the reallocation of employment towards more interactive tasks is larger in metro than in non-metro areas.

We organize our empirical analysis around a simple model of agglomeration that incorporates specialization across locations, occupations and sectors. These rich dimensions of specialization enable the model to account for the observed changes in the tasks undertaken in metro and non-metro areas over time. First, the model explains the reallocation of employment over time towards more interactive occupations in terms of differences in productivity growth across occupations and inelastic demand. Second, the model predicts that this reallocation is stronger in metro areas than in non-metro areas because of falling trade costs and specialization according to comparative advantage. Third, and for the same reason, the model predicts that the reallocation towards more interactive occupations is stronger in denser metro areas. Fourth, the model explains how this allocation of tasks across areas that differ in their population density could have taken place both between sectors and within sectors. Fifth, the model accounts for our finding that in 1880, when trade in tasks was still costly, metro areas were not more interactive than non-metro areas. Finally, the model explains faster population growth over time in metro than in non-metro areas in terms of more rapid productivity growth in metro areas.

While theories of agglomeration have traditionally emphasized the movement of goods and people, our findings are consistent with the view that the generation, communication and exchange of ideas are increasingly important for agglomeration. Although we have used our methodology to examine task specialization in urban and rural areas, it also has a wide range of potential other applications. For example, our approach could be used to examine the extent to which occupations require narrow versus broad ranges of tasks, or it could be used to construct employment and wagebill shares for individual tasks and examine the extent to which they are related to trade and technology.

## References

- [1] Alonso W (1964) *Location and Land Use*, Cambridge, MA: Harvard University Press.
- [2] Autor, David H. and David Dorn (2013) “The Growth of Low Skill Service Jobs and the Polarization of the U.S. Labor Market,” *American Economic Review*, forthcoming.
- [3] Autor, David H. and Michael J. Handel (2009) “Putting Tasks to the Test: Human Capital, Job Tasks and Wages,” *NBER Working Paper*, 15116.
- [4] Autor, David H., Frank Levy and Richard J. Murnane (2003) “The Skill Content of Recent Technological Change: An Empirical Exploration,” *Quarterly Journal of Economics*, November, 1279-1333.
- [5] Bacolod, Marigee, Bernardo S. Blum and William C. Strange (2009a) “Skills in the City,” *Journal of Urban Economics*, 65, 136-153.
- [6] Bacolod, Marigee, Bernardo S. Blum and William C. Strange (2009b) “Urban Interactions: Soft Skills Versus Specialization,” *Journal of Economic Geography*, 9, 227-262.
- [7] Baldwin, Richard E. (2012) “Trade and Industrialisation after Globalisation’s 2nd Unbundling: How Building and Joining a Supply Chain are Different and Why it Matters,” *CEPR Discussion Paper*, 8768.
- [8] Baumgardner, J. (1988) “The Division of Labor, Local Markets, and Worker Organization,” *Journal of Political Economy*, 96, 509-527.
- [9] Becker, Gary S. and Kevin M. Murphy (1992) “The Division of Labor, Coordination Costs, and Knowledge,” *Quarterly Journal of Economics*, 107(4), 1137-1160.
- [10] Becker, Sascha O., Karolina Ekholm and Marc-Andreas Muendler (2009) “Offshoring and the Onshore Composition of Tasks and Skills,” University of California, San Diego.
- [11] Blinder, Alan (2009) “How Many U.S. Jobs Might Be Offshorable,” *World Economics*, 10(2), 41-78.
- [12] Blinder, Alan and Alan Krueger (2012) “Alternative Measures of Offshorability: A Survey Approach,” *Journal of Labor Economics*, forthcoming.
- [13] Costa, Dora L. and Matthew E. Kahn (2000) “Power Couples: Changes In The Locational Choice Of The College Educated, 1940-1990,” *Quarterly Journal of Economics*, 115(4), 1287-1315.
- [14] Costinot, A., D. Donaldson and I. Komunjer (2012) “What Goods do Countries Trade? A Quantitative Exploration of Ricardo’s Ideas,” *Review of Economic Studies*, 79(2), 581-608.
- [15] Davis, Donald R. and Jonathan I. Dingel (2012) “A Spatial Knowledge Economy,” *NBER Working Paper*, 18188.
- [16] Davis, Donald R. and Jonathan I. Dingel (2013) “The Comparative Advantage of Cities,” Columbia University, mimeograph.
- [17] Davis, Morris A. and François Ortalo-Magne (2011) “Household Expenditures, Wages, Rents,” *Review of Economic Dynamics*, 14(2), 248-261.

- [18] Duranton, Gilles (1998) "Labor Specialization, Transports Costs and City Size," *Journal of Regional Science*, 38 (4), 553-573.
- [19] Duranton, Gilles and Diego Puga (2004) "Micro-foundations of Urban Agglomeration Economies," Chapter 48 in (eds.) J. V. Henderson and J.F. Thisse, *Handbook of Regional and Urban Economics*, Volume 4, Amsterdam: Elsevier North Holland.
- [20] Duranton, Gilles and Diego Puga (2005) "From Sectoral to Functional Urban Specialization," *Journal of Urban Economics*, 343-370.
- [21] Duranton, Gilles and Hubert Jayet (2011) "Is the Division of Labour Limited by the Extent of the Market? Evidence from French Cities," *Journal of Urban Economics*, 69(1), 56-71.
- [22] Eaton, Jonathan and Samuel S. Kortum (2002) "Technology, Geography, and Trade," *Econometrica*, 70(5) (2002), 1741-1779.
- [23] Ellison, Glenn, Edward Glaeser and William R. Kerr (2010) "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," *American Economic Review*, 100(3), 1195-1213.
- [24] Ethier, Wilfred J. (1982) "National and International Returns to Scale in the Modern Theory of International Trade," *American Economic Review*, 72(3), 389-405.
- [25] Fujita, Masahisa, Paul Krugman and Anthony J. Venables (1999) *The Spatial Economy: Cities, Regions and International Trade*, Cambridge: MIT Press.
- [26] Fujita, Masahisa and Takatoshi Tabuchi (1997) "Regional Growth in Postwar Japan," *Regional Science and Urban Economics*, 27, 643-670.
- [27] Gaspar, Jess and Edward L. Glaeser (1998) "Information Technology and the Future of Cities," *Journal of Urban Economics*, 43(1), 136-156.
- [28] Gentzkow, Matthew, Jesse M. Shapiro and Michael Sinkinson (2012) "Competition and Ideological Diversity: Historical Evidence from U.S. Newspapers," Chicago Booth School of Business, mimeograph.
- [29] Glaeser, Edward L. and Janet E. Kohlhase (2003) "Cities, Regions and the Decline of Transport Costs," *Papers in Regional Science*, 83(1), 197-228.
- [30] Glaeser, Edward L., A. M. Ponzetto and Kristina Toblo (2011) "Cities, Skills and Regional Change," *NBER Working Paper*, 16934.
- [31] Glaeser, Edward L. and Matthew G. Resseger (2009) "The Complementarity Between Cities and Skills," *NBER Working Paper*, 15103.
- [32] Glaeser, Edward L. and Albert Saiz (2003) "The Rise of the Skilled City," *NBER Working Papers*, 10191.
- [33] Gray, Rowena (2010) "Taking Technology to Task: The Skill Content of Technological Change in Early Twentieth Century U.S." University of Essex, mimeograph.
- [34] Grossman, Gene and Esteban Rossi-Hansberg (2008) "Trading Tasks: A Simple Theory of Offshoring," *American Economic Review*, 98(5), 1978-1997.

- [35] Grossman, G. M. and E. Rossi-Hansberg (2012) “Task Trade Between Similar Countries,” *Econometrica*, 80(2), 593-629.
- [36] Helpman, Elhanan (1998) “The Size of Regions,” In *Topics in Public Economics: Theoretical and Applied Analysis*, ed. David Pines, Efraim Sadka, and Itzhak Zilcha, Cambridge: Cambridge University Press.
- [37] Helsley, Robert W. and William C. Strange (2007) “Agglomeration, Opportunism and the Organization of Production,” *Journal of Urban Economics*, 62, 55-75.
- [38] Henderson, J. Vernon (1974) “The Sizes and Types of Cities,” *American Economic Review*, 64(4), 640-656.
- [39] Hüllen, Werner (2003) *A History of Roget’s Thesaurus: Origins, Development, and Design*, Oxford: Oxford University Press.
- [40] Lin, Jeffrey (2011) “Technological Adaptation, Cities and New Work,” *Review of Economics and Statistics*, 93(2), 554-574.
- [41] Michaels, Guy, Ferdinand Rauch and Stephen Redding (2012) “Urbanization and Structural Transformation,” *Quarterly Journal of Economics*, 127, 535-586.
- [42] Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, and Erez Lieberman Aiden (2011) “Quantitative Analysis of Culture Using Millions of Digitized Books,” *Science*, 331, 176-182.
- [43] Mills, E. S. (1967) “An Aggregative Model of Resource Allocation in a Metropolitan Area,” *American Economic Review*, 57, 197-210.
- [44] Moretti, Enrico (2004) “Human Capital Externalities in Cities,” Chapter 51 in (eds) V. Henderson and J.F. Thisse, *Handbook of Urban and Regional Economics*, Amsterdam: North Holland-Elsevier.
- [45] Muth, R. (1968) *Cities and Housing*, Chicago: University of Chicago Press.
- [46] Ngai, Rachel, and Chris Pissarides (2007) “Structural Change in a Multi-sector Model of Growth,” *American Economic Review*, 97(1), 429-443.
- [47] Ota, Mitsuru and Masahisa Fujita (1993) “Communication Technologies and Spatial Organization of Multi-unit Firms in Metropolitan Areas,” *Regional Science and Urban Economics*, 23, 695-729.
- [48] Ottaviano, Gianmarco I. P., Giovanni Peri and Greg C. Wright (2010) “Immigration, Offshoring and American Jobs,” *NBER Working Paper*, 16439.
- [49] Roget, Peter Mark (1911) *Thesaurus of English Words and Phrases*, Computer Readable Version of 1911 Edition, Evergreen Review, Inc.
- [50] Rosenthal, Stuart S. and William C. Strange (2004) “Evidence on the Nature and Sources of Agglomeration Economics,” Chapter 49 in (eds.) J. V. Henderson and J.F. Thisse, *Handbook of Regional and Urban Economics*, Volume 4, Amsterdam: Elsevier North Holland.
- [51] Rossi-Hansberg, Esteban, Pierre-Daniel Sarte and Raymond Owens III (2009) “Firm Fragmentation and Urban Patterns,” *International Economic Review*, 50(1), 143-186.

- [52] Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek (2010) *Integrated Public Use Microdata Series*, Version 5.0 [Machine-readable database]. Minneapolis: University of Minnesota.
- [53] U. S. Department of Labor (1939) *Dictionary of Occupational Titles*, Washington DC: U.S. Department of Labor.
- [54] U. S. Department of Labor (1991) *Dictionary of Occupational Titles*, Washington DC: U.S. Department of Labor.
- [55] Yi, Kei-Mu and Jing Zhang (2010) “Structural Change in an Open Economy,” University of Minnesota, mimeograph.

**Table 1: Metro Area Specialization for Aggregate Occupations and Sectors**

<b>Panel A</b>						
<b>Two-digit occupation</b>	<b>Coefficient</b>	<b>Standard</b>		<b>Coefficient</b>	<b>Standard</b>	
	<b>1880</b>	<b>Error 1880</b>	<b>Rank 1880</b>	<b>2000</b>	<b>Error 2000</b>	<b>Rank 2000</b>
Clerical and Kindred	0.15	0.08	1	0.04	0.01	4
Craftsmen	0.09	0.06	2	-0.01	0.01	6
Operatives	0.06	0.07	3	-0.05	0.01	7
Sales workers	0.01	0.07	4	0.05	0.01	2
Service Workers	0.00	0.08	5	0.00	0.01	5
Managers, Officials, and Proprietors	-0.03	0.08	6	0.05	0.01	3
Professional, Technical	-0.07	0.08	7	0.07	0.01	1
Laborers	-0.20	0.18	8	-0.15	0.07	8

<b>Panel B</b>						
<b>Two-digit sector</b>	<b>Coefficient</b>	<b>Standard</b>		<b>Coefficient</b>	<b>Standard</b>	
	<b>1880</b>	<b>Error 1880</b>	<b>Rank 1880</b>	<b>2000</b>	<b>Error 2000</b>	<b>Rank 2000</b>
Entertainment and Recreation Services	0.29	0.08	1	0.04	0.01	4
Wholesale and Retail Trade	0.13	0.05	2	0.01	0.01	6
Finance, Insurance, and Real Estate	0.13	0.06	3	0.06	0.01	2
Manufacturing	0.06	0.05	4	-0.01	0.01	10
Personal Services	0.01	0.06	5	0.03	0.01	5
Transportation, Communication, Other Utilities	0.01	0.04	6	0.05	0.01	3
Public Administration	-0.03	0.07	7	0.01	0.01	7
Professional and Related Services	-0.03	0.06	8	0.00	0.01	9
Business and Repair Services	-0.12	0.08	9	0.08	0.01	1
Construction	-0.14	0.08	10	0.00	0.01	8
Mining	-0.31	0.05	11	-0.27	0.03	11

Notes: Coefficients estimated from a regression of the share of employment in metro areas in an occupation-sector-year on occupation-year and sector-year fixed effects (regression (11) in the paper). Occupation-year and sector-year fixed effects are each normalized to sum to zero. A separate regression is estimated for each year. Standard errors are clustered by occupation. Occupations and sectors sorted by the rank of their estimated coefficients for 1880.

**Table 2: Verbs Most and Least Strongly Correlated with Metro Area Employment Shares**

<b>Panel A: Verbs Most Strongly Correlated with Metro Area Employment Shares</b>							
<b>Rank</b>	<b>1880</b>	<b>1900</b>	<b>1920</b>	<b>1940</b>	<b>1960</b>	<b>1980</b>	<b>2000</b>
1	Thread	Thread	File	File	Document	Identify	Develop
2	Stretch	Stitch	Distribute	Bill	Schedule	Document	Determine
3	Interfere	Telephone	Record	Take	File	Advise	Analyze
4	Hand	Sew	Notice	Compile	Record	Concern	Factor
5	Ravel	Hand	Telephone	Distribute	Distribute	Report	Review
6	Sew	Assist	Bill	Pay	Compile	Schedule	Confer
7	Braid	Visit	Envelope	Letter	Notice	Develop	Advise
8	Visit	Describe	Document	Notice	Identify	Analyze	Report
9	Receive	Number	Learn	Record	Send	Determine	Concern
10	Sack	Stamp	Number	Send	Notify	Notify	Plan
<b>Panel B: Verbs Least Strongly Correlated with Metro Area Employment Shares</b>							
<b>Rank</b>	<b>1880</b>	<b>1900</b>	<b>1920</b>	<b>1940</b>	<b>1960</b>	<b>1980</b>	<b>2000</b>
1821	Conduct	Abstract	Counsel	Delegate	Accord	Power	Restrain
1822	Teach	Tread	Discuss	Enlist	Feed	Pour	Cut
1823	Channel	Pinch	Hear	Labor	Escape	Erect	Power
1824	Sound	Assign	Assign	Tread	Hook	Clean	Massage
1825	Rule	Settle	Teach	Assign	Traverse	Massage	Remove
1826	Matter	Matter	Matter	Approve	Tread	Pump	Feed
1827	Drill	Tunnel	Consolidate	Extract	Loosen	Cut	Clean
1828	Tread	Sound	Rule	Tunnel	Range	Feed	Pump
1829	Tunnel	Rule	Tunnel	Malt	Activate	Move	Move
1830	Pinch	Sole	Sound	Establish	Turn	Turn	Turn

Notes: Coefficients estimated from a regression of the share of occupation-sector employment in metro areas on the frequency with which a verb is used for an occupation and verb-sector-year fixed effects (regression (12) in the paper). A separate regression is estimated for each verb and verbs are sorted by their estimated coefficients normalized by the standard deviation for the verb frequency. Verbs are from the time-invariant occupational descriptions from the 1991 Dictionary of Occupations (DOTs).

Table 3: Correlation of Thesaurus Sections with Metro Area Employment Shares

Thesaurus Class	Thesaurus Section	Rank	Rank	Rank
		Section 1880	Section 2000	1880 - 2000
C1, Abstract relations	SECTION I. EXISTENCE	9	10	-1
C1, Abstract relations	SECTION II. RELATION	17	6	11
<b><i>C1, Abstract relations</i></b>	<b><i>SECTION III. QUANTITY</i></b>	<b><i>1</i></b>	<b><i>32</i></b>	<b><i>-31</i></b>
C1, Abstract relations	SECTION IV. ORDER	30	11	19
C1, Abstract relations	SECTION V. NUMBER	23	14	9
<b><i>C1, Abstract relations</i></b>	<b><i>SECTION VI. TIME</i></b>	<b><i>2</i></b>	<b><i>20</i></b>	<b><i>-18</i></b>
C1, Abstract relations	SECTION VII. CHANGE	36	7	29
C1, Abstract relations	SECTION VIII. CAUSATION	28	21	7
C2, Space	SECTION I. SPACE I N GENERAL	8	33	-25
<b><i>C2, Space</i></b>	<b><i>SECTION II. DIMENSIONS</i></b>	<b><i>4</i></b>	<b><i>36</i></b>	<b><i>-32</i></b>
C2, Space	SECTION IV. MOTION	25	24	1
<b><i>C3, Matter</i></b>	<b><i>SECTION I. MATTER IN GENERAL</i></b>	<b><i>3</i></b>	<b><i>31</i></b>	<b><i>-28</i></b>
<b><i>C3, Matter</i></b>	<b><i>SECTION II. INORGANIC MATTER</i></b>	<b><i>5</i></b>	<b><i>35</i></b>	<b><i>-30</i></b>
C3, Matter	SECTION III. ORGANIC MATTER	14	37	-23
C4, Intellect	SECTION I. NATURE OF IDEAS COMMUNICATED	19	15	4
C4, Intellect	SECTION I. OPERATIONS OF INTELLECT IN GENERAL	15	23	-8
C4, Intellect	SECTION II. MODES OF COMMUNICATION	18	9	9
<b><i>C4, Intellect</i></b>	<b><i>SECTION II. PRECURSORY CONDITIONS &amp; OPERATIONS</i></b>	<b><i>34</i></b>	<b><i>5</i></b>	<b><i>29</i></b>
C4, Intellect	SECTION III. MATERIALS FOR REASONING	33	8	25
<b><i>C4, Intellect</i></b>	<b><i>SECTION III. MEANS OF COMMUNICATING IDEAS</i></b>	<b><i>11</i></b>	<b><i>2</i></b>	<b><i>9</i></b>
C4, Intellect	SECTION IV. REASONING PROCESSES	38	22	16
<b><i>C4, Intellect</i></b>	<b><i>SECTION V. RESULTS OF REASONING</i></b>	<b><i>7</i></b>	<b><i>1</i></b>	<b><i>6</i></b>
C4, Intellect	SECTION VI. EXTENSION OF THOUGHT	24	18	6
C4, Intellect	SECTION VII. CREATIVE THOUGHT	35	17	18
C5, Will	SECTION I. GENERAL INTERSOCIAL VOLITION	20	25	-5
C5, Will	SECTION I. VOLITION IN GENERAL	29	28	1
C5, Will	SECTION II. Prospective Volition 1	21	38	-17
C5, Will	SECTION II. SPECIAL INTERSOCIAL VOLITION	26	13	13
C5, Will	SECTION III. CONDITIONAL INTERSOCIAL VOLITION	27	12	15
<b><i>C5, Will</i></b>	<b><i>SECTION III. VOLUNTARY ACTION</i></b>	<b><i>32</i></b>	<b><i>4</i></b>	<b><i>28</i></b>
C5, Will	SECTION IV. ANTAGONISM	10	27	-17
C5, Will	SECTION IV. POSSESSIVE RELATIONS	16	16	0
C5, Will	SECTION V. RESULTS OF VOLUNTARY ACTION	31	26	5
C6, Emotion, Religion, Morality	SECTION I. AFFECTIONS IN GENERAL	6	34	-28
C6, Emotion, Religion, Morality	SECTION II. PERSONAL AFFECTIONS	22	30	-8
C6, Emotion, Religion, Morality	SECTION III. SYMPATHETIC AFFECTIONS	12	29	-17
<b><i>C6, Emotion, Religion, Morality</i></b>	<b><i>SECTION IV. MORAL AFFECTIONS</i></b>	<b><i>37</i></b>	<b><i>3</i></b>	<b><i>34</i></b>
C6, Emotion, Religion, Morality	SECTION V. RELIGIOUS AFFECTIONS	13	19	-6

Notes: Coefficients estimated from a regression of the share of occupation-sector employment in metro areas on the frequency with which the verbs used for an occupation are classified within thesaurus sections and thesaurus-section-sector-year fixed effects (regression (14) in the paper). A separate regression is estimated for each thesaurus section. Verbs are from the time-invariant occupational descriptions from the 1991 Dictionary of Occupations (DOTs). Thesaurus sections ranked in terms of their estimated coefficient normalized by the standard deviation for the thesaurus section frequency, where the largest value is assigned a rank of one. Top-five thesaurus sections in 1880 highlighted in bold and italics. Top-five thesaurus sections in 2000 highlighted in bold.

**Table 4: Most and Least Interactive Occupations**

**Panel A: Top Ten Interactive Occupations**

Economists  
Nurses, professional  
Pharmacists  
Clergymen  
Religious workers  
Accountants and auditors  
Postmasters  
Buyers and dept heads, store  
Aeronautical-Engineers  
Statisticians and actuaries

**Panel B: Bottom Ten Interactive Occupations**

Brickmasons,stonemasons, tile setters  
Attendants, auto service and parking  
Painters, except construction  
Plumbers and pipe fitters  
Upholsterers  
Asbestos and insulation workers  
Welders and flame cutters  
Blasters and powdermen  
Dressmakers and seamstresses  
Roofers and slaters

Notes: The table reports the ten occupations with the lowest and highest interactiveness, as measured by the frequency of verb use in Classes IV and V of Roget's Thesaurus. Verbs are from the time-invariant occupational descriptions from the 1991 Dictionary of Occupations (DOTs).

**Table 5: Metro Employment and Wagebill Shares and Interactiveness**

<b>Panel A: Between sectors</b>							
<b>Measure</b>	<b>1880</b>	<b>1900</b>	<b>1920</b>	<b>1940</b>	<b>1960</b>	<b>1980</b>	<b>2000</b>
Interactiveness (Employment)	-0.130 (0.267)	-0.132 (0.239)	0.258 (0.419)	0.556 (0.405)	0.728*** (0.267)	0.901*** (0.200)	0.814*** (0.182)
Thought (Employment)	-0.649** (0.268)	-1.304*** (0.261)	-1.805*** (0.363)	-0.608 (0.493)	0.179 (0.313)	0.780*** (0.280)	1.202*** (0.237)
Communication (Employment)	-0.412*** (0.153)	-0.568*** (0.153)	-0.641*** (0.188)	-0.212 (0.272)	0.219 (0.199)	0.359* (0.210)	0.530** (0.233)
Intersocial (Employment)	-0.292** (0.144)	-0.473*** (0.136)	-0.548*** (0.169)	-0.0624 (0.203)	0.126 (0.133)	0.280** (0.124)	0.342*** (0.109)
Interactiveness (Wage Bill)				0.557 (0.366)	0.557* (0.283)	0.814*** (0.215)	0.733*** (0.201)
<b>Panel B: Within sectors</b>							
<b>Measure</b>	<b>1880</b>	<b>1900</b>	<b>1920</b>	<b>1940</b>	<b>1960</b>	<b>1980</b>	<b>2000</b>
Interactiveness (Employment)	-0.410*** (0.120)	-0.261** (0.119)	-0.104 (0.119)	-0.0360 (0.119)	0.190*** (0.0644)	0.274*** (0.0514)	0.317*** (0.0402)
Thought (Employment)	-0.340** (0.134)	-0.411*** (0.132)	-0.299*** (0.0933)	-0.145 (0.0948)	0.153*** (0.0489)	0.227*** (0.0374)	0.246*** (0.0394)
Communication (Employment)	-0.0408 (0.144)	-0.0423 (0.118)	0.0249 (0.0977)	0.118 (0.0789)	0.183*** (0.0360)	0.168*** (0.0323)	0.140*** (0.0384)
Intersocial (Employment)	-0.0300 (0.130)	-0.0809 (0.0780)	-0.0172 (0.0582)	0.0197 (0.0492)	0.105*** (0.0320)	0.0652* (0.0342)	0.0460 (0.0476)
Inteactiveness (Wage Bill)				0.043 (0.087)	0.207*** (0.053)	0.281*** (0.043)	0.311*** (0.037)
Sector-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes

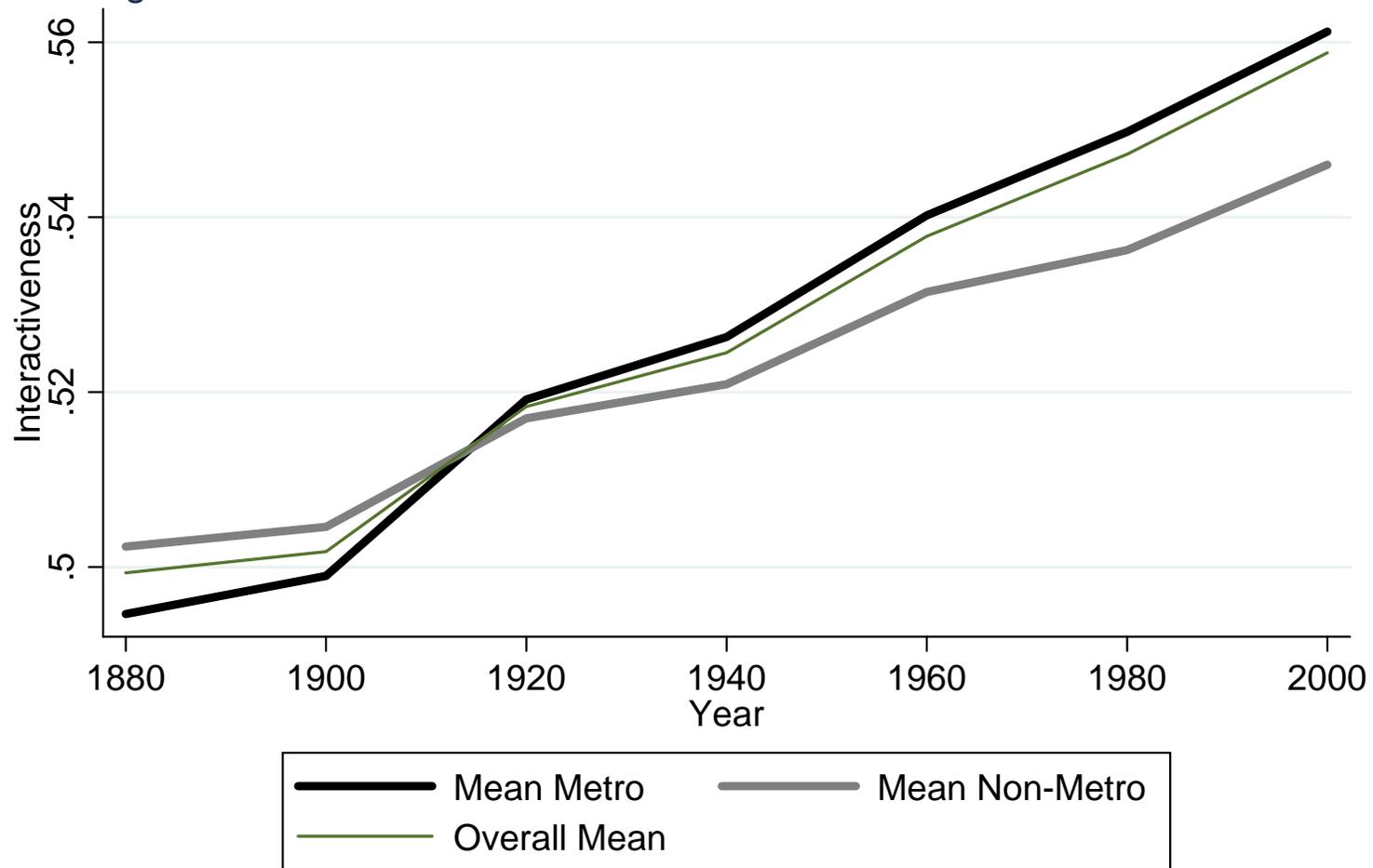
Notes: Each cell of each panel of the table corresponds to a separate regression. The left-hand side in the first four rows of each panel is the share of employment in metro areas; the left-hand side in the fifth row of each panel is the share of the wagebill in metro areas; the wagebill data are only available from 1940 onwards; Interactiveness is our baseline measure using the 1991 DOTs and Classes IV-V of the thesaurus; Thought uses Class IV (Division 1) of the thesaurus; Communication uses Class IV (Division 2) of the thesaurus; Intersocial uses Class V (Division 2) of the thesaurus. In Panel A, the sample is a cross-section of three-digit sectors for each year, and the standard errors are heteroscedasticity robust. In Panel B, the sample is a panel of sectors and occupations for each year; sector-year fixed effects are included; and the standard errors are clustered on occupation. See Section 8.2 for further details on the estimated equation. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.

**Table 6: Metro Area Employment Shares and Interactiveness, Within-sector and Within-Occupation**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Metro	Metro	Metro	Metro	Metro	Metro	Metro	Metro	Metro
	Employment	Employment	Employment	Employment	Employment	Employment	Employment	Employment	Employment
	Share	Share	Share	Share	Share	Share	Share	Share	Share
Interactiveness x 1900	0.00207 (0.141)	0.104 (0.162)	0.0493 (0.119)	0.292 (0.195)	0.108 (0.178)	0.124 (0.177)	0.402 (0.246)	-0.0307 (0.0763)	0.124 (0.146)
Interactiveness x 1920	0.186 (0.202)	0.187 (0.218)	0.132 (0.176)	0.525** (0.223)	0.254 (0.250)	0.272 (0.203)	0.541* (0.277)	-0.00455 (0.129)	0.321 (0.206)
Interactiveness x 1940	0.399 (0.243)	0.321 (0.235)	0.287 (0.205)	0.455* (0.254)	0.334 (0.236)	0.324 (0.233)	0.732** (0.290)	0.0379 (0.117)	0.369* (0.221)
Interactiveness x 1960	0.573** (0.231)	0.485*** (0.185)	0.316** (0.158)	0.548** (0.261)	0.284 (0.228)	0.410* (0.227)	0.842*** (0.273)		
Interactiveness x 1980	0.677*** (0.244)	0.560*** (0.174)	0.489*** (0.152)	0.627** (0.258)	0.424* (0.250)	0.515** (0.240)	0.893*** (0.268)	0.233*** (0.0672)	0.595*** (0.204)
Interactiveness x 2000	0.672*** (0.253)	0.596*** (0.174)	0.609*** (0.141)	0.788*** (0.233)	0.552** (0.276)	0.681*** (0.221)	0.874*** (0.275)	0.261*** (0.0670)	0.823*** (0.167)
Observations	56,760	56,760	49,108	41,442	25,105	30,593	48,210	35,662	44,128
Year fixed effects	yes								
Sector fixed effects	yes								
Occupation fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes
Sector-Year fixed effects		yes	yes	yes	yes	yes	yes	yes	yes
Married only sample			yes						
Single only sample				yes					
Manufacturing only sample					yes				
Services only sample						yes			
Excluding headquarters occupations							yes		
Excluding more skilled metro areas								yes	
Excluding less skilled metro areas									yes

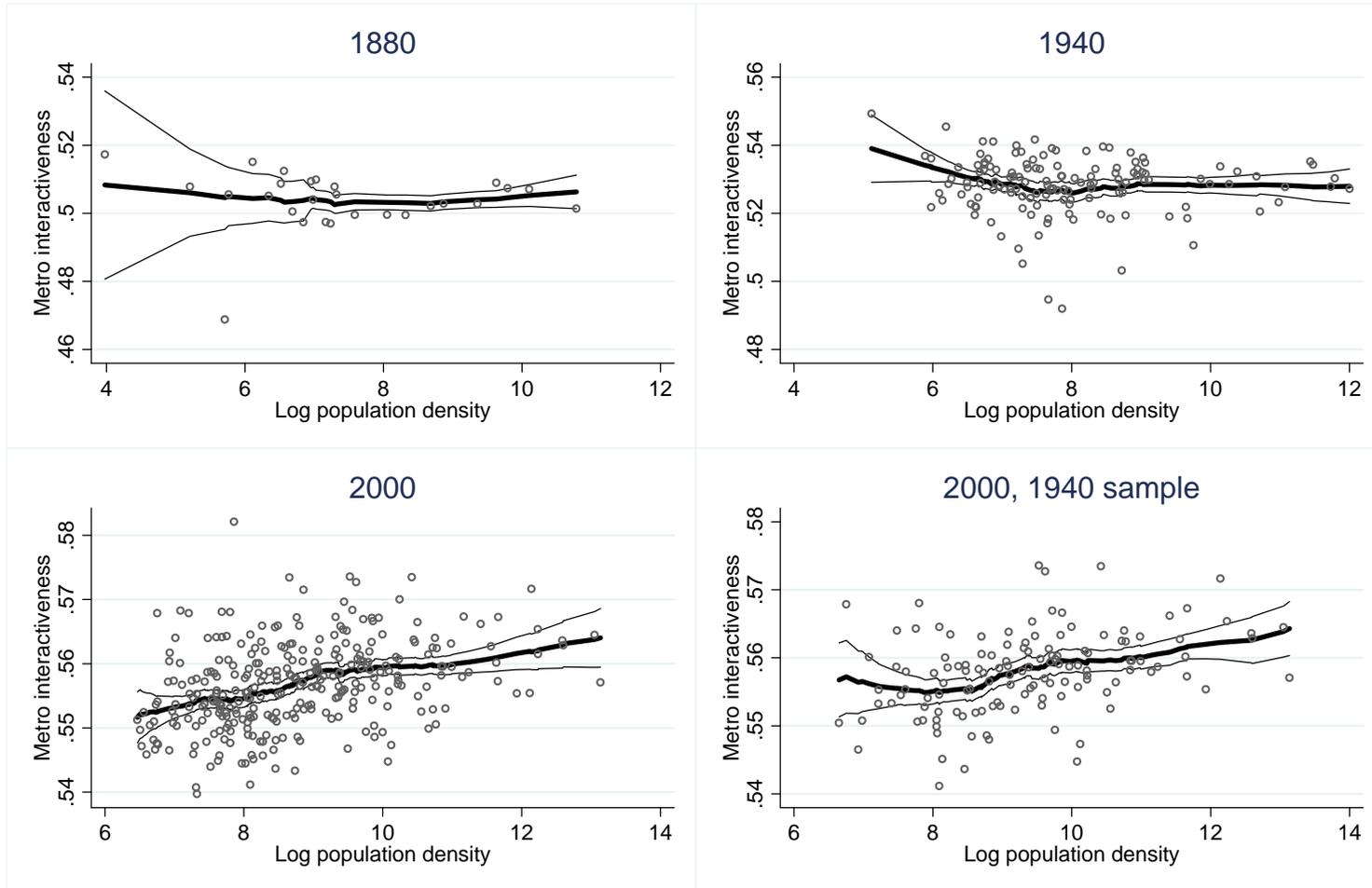
Notes: Sample is a panel of occupation-sector-year observations for twenty-year intervals from 1880-2000; 1880 is the excluded year from the interactions; interactiveness is our baseline measure using the 1991 DOTs and Classes IV-V of the thesaurus. Married only sample includes married workers only. Single only sample excludes married workers. Manufacturing only sample includes workers in manufacturing only. Services only sample includes workers in services only. Sample excluding headquarters occupations excludes 22 three-digit occupations typically concentrated in headquarters. More and less-skilled metro areas are defined as in Glaeser and Resseger (2009) based on whether the share of adults with a college degree in a Metropolitan Statistical Area (MSA) is greater than or less than 25.025 percent in 2006. The year 1960 is omitted in Columns (8) and (9) because the IPUMS 1960 data do not contain the identifiers for individual MSAs. See Section 8.2 for further details on the estimated equation. Standard errors are clustered on occupation; \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.

Figure 1: Mean Interactiveness in Metro and Non-Metro Areas over Time



Notes: Mean interactiveness computed using time-invariant occupational descriptions from the 1991 DOTs.

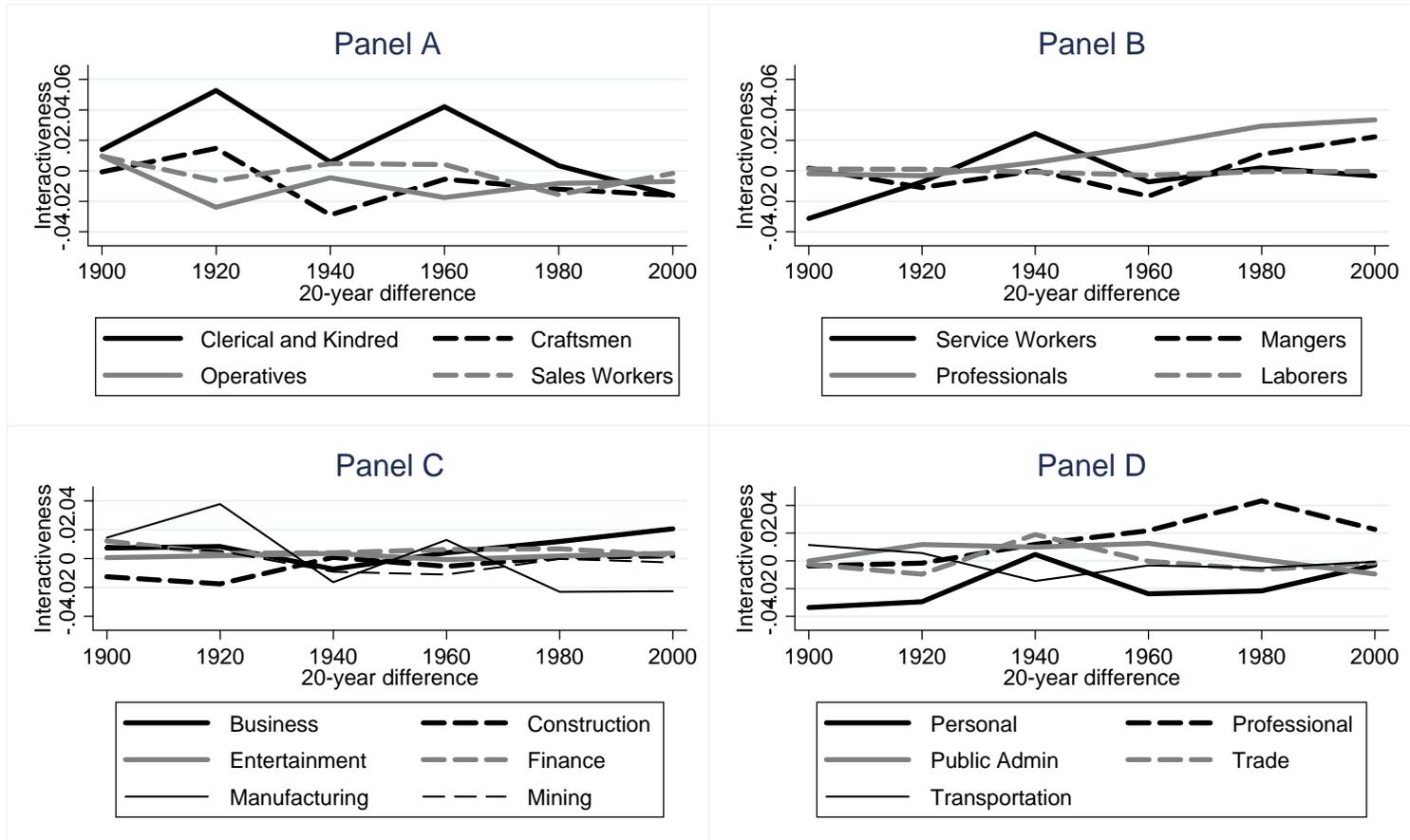
Figure 2: Mean Interactiveness Across Metro Areas



Notes: Mean interactiveness computed using time-invariant occupational descriptions from the 1991 DOTs. Thick solid line is fitted values from locally-weighted linear least squares regression. Thin solid lines are 95 percent point confidence intervals.

STATA™

### Figure 3: Decomposition of Difference in Change in Interactiveness Between Metro and Non-Metro Areas



Notes: Decomposition of the difference between metro and non-metro areas in the change in mean interactiveness over twenty-year intervals (equation (19) in the paper) into the contributions of two-digit occupations and sectors. Mean interactiveness based on time-invariant occupational descriptions from the 1991 DOTs.