The Division of Labor, Coordination, and the Demand for Information Processing

Guy Michaels*

This Version: June 2007

Abstract

Since Adam Smith's time, the division of labor in production has increased significantly, while information processing has become an important part of work. This paper examines whether the need to coordinate an increasingly complex division of labor has raised the demand for clerical office workers, who process information that is used to coordinate production. In order to examine this question empirically, I introduce a measure of the complexity of an industry's division of labor that uses the Herfindahl index of occupations it employs, excluding clerks and managers. Using US data I find that throughout the 20th century more complex industries employed relatively more clerks, and recent Mexican data shows a similar relationship. The relative complexity of industries is persistent over time and correlated across these two countries. I further document the relationship between complexity and the employment of clerks using an early information technology (IT) revolution that took place around 1900, when telephones, typewriters, and improved filing techniques were introduced. This IT revolution raised the demand for clerks in all manufacturing industries, but significantly more so in industries with a more complex division of labor. Interestingly, recent reductions in the price of IT have enabled firms to substitute computers for clerks, and I find that more complex industries have substituted clerks more rapidly.

Keywords: Information Processing, Division of Labor, Technological Change, Organization of Production.

JEL classification: J44, M54, D73, O33.

---

*Centre for Economic Performance and Department of Economics, LSE, Houghton Street, London WC2A 2AE, UK. E-mail: g.michaels [at] lse.ac.uk. An earlier version of this paper from April 2004 was entitled: "Technology, Complexity and Information: the Evolution of Demand for Office Workers." I am grateful to Daron Acemoglu and Joshua Angrist for their many helpful suggestions. I also thank David Autor, Dora Costa, Luis Garicano, Claudia Goldin, Todd Gormley, Alan Manning, Gerard Padro-i-Miquel, Michael Piore, Andrea Prat, Peter Temin, Manuel Trajtenberg, John Van Reenen, Alex Whalley and seminar and workshop participants at Harvard, LSE, MIT, and SOLE 2006.

1 Introduction

Since Adam Smith's analysis of the pin factory (1776), the division of labor in production has become much more detailed. Today, the production of most goods involves the coordination of a wide range of occupations. At the same time, white collar work that involves information processing and coordination has become important in all firms and organizations (Autor, Levy, and Murnane 2003). This paper examines the extent to which a more complex division of labor entails higher costs of processing the information that is used to coordinate production.

Most of the literature on coordination costs and the division of labor is theoretical. Radner (1992, 1993), Garicano (2000), and Garicano and Rossi-Hansberg (2004) consider workers' capacity for processing information and the formation of knowledge hierarchies. In "The Division of Labor, Coordination, and Knowledge" (1992), Becker and Murphy argue that the division of labor is constrained by the costs of coordinating specialized workers and by the availability of information. But it has proved difficult to estimate these costs of coordination, because of the difficulty of measuring the division of labor empirically.[1]

In this paper, I characterize a production process as complex when it requires a diverse set of occupations. Specifically, I define the complexity of a manufacturing industry as one minus the Herfindahl index of the occupations of its employees, excluding managers, clerks, accountants, and auditors.[2] For example, textile and clothing industries employ an occupationally homogenous workforce and therefore have low complexity. By contrast, manufacturers of transportation equipment employ an occupationally heterogeneous workforce, so their complexity is high.

How does the complexity of the division of labor affect the demand for information processing? Consider two factories that employ an equal number of workers; in one factory all the workers perform the same task, while in the other each worker performs a different task.

[1] Recent work by Garcano and Hubbard (2003), analyzes the division of labor among lawyers, and Duranton and Jayet (2006) find that specialized occupations are more prevalent in large cities.

[2] Most of the analysis uses fixed 1950 census definitions of industries and occupations.

In order to effectively coordinate production in the more complex factory, labor has to be allocated across tasks, different complementary inputs need to be acquired and distributed, and various inventories need to be tracked. Thus, an effective coordination of a more complex labor force requires more information.

In order to examine the relationship between complexity and demand for information processing, I consider the employment of clerical office workers (clerks). Clerks are skilled workers, who use information technology (IT) equipment to generate, store, and communicate information. The information that clerks process is an intermediate input that facilitates the coordination of production. Although this paper focuses on the demand for clerks, it also sheds light on the demand for managers, who analyze the information produced by clerks and use it to coordinate production.

Production of manufactured goods in US industries has become much more complex from 1860-2000. Over this period, the average complexity of U.S. manufacturing industries has increased from about 0.4 to 0.78. At the same time, the fraction of clerks in manufacturing has risen from less than 1 percent to over 11 percent. Although this relationship is suggestive, many other factors have changed over this period, so I turn to cross-industry comparisons of complexity and the fraction of clerks employed.

Figure 1 shows that throughout the 20th century more complex industries employed more clerks. This correlation strengthened considerably over the first half of the 20 century. In recent decades it appears to have weakened, but in 2000 it was still statistically significant.

The positive correlation between complexity and the demand for clerks is consistent with the hypothesis that coordination is costly, but it leaves the possibility of reverse causality: can industries with more clerks afford greater complexity? It is especially difficult to address this question, since the relative complexity industries is stable over time, likely reflecting persistent differences in their production technologies. In absence of exogenous variation in the level of complexity, I examine whether more complex industries had a latent demand for information even before it was cost-effective to process information.

Before 1880 IT equipment was very limited, and so was the use of clerks. But over the next 30 years telephones, typewriters, and vertical filing techniques came into widespread use (see Figure 2). As a result of this early IT Revolution, firms started to employ clerks in tandem with IT equipment to produce information that was used to coordinate production. Figures 3 and 4 show that the fraction of clerks in manufacturing increased from about 1 percent to over 6 percent from 1880-1910, while their wages relative to production workers remained stable.[3] This evidence suggests that this early IT revolution increased the relative demand for clerks. More importantly, I find that industries that were more complex in 1880 increased their demand for clerks significantly more than other industries over the next 30 years.

Once firms use information, a further decrease in the effective price of IT equipment may cause them to substitute it for clerks. This is indeed what happened from 1960-2000, when the employment share and relative wages of clerks decreased (Figures 3 and 4). This finding provides a concrete case in which technology and a specific set of skills are substitutes rather than complements. Interestingly, I find that complex industries increased their demand for computer personnel (likely correlated with their use of IT equipment) and decreased their demand for clerks more rapidly than other industries.

In order to further test the hypothesis that differences in demand for clerks are driven by underlying technological differences, I compute measures of complexity using 1990 classification of occupations for both the US and Mexico. Despite differences in the classification of occupations across the two countries, I find that the cross-industry correlation of complexity is about 0.3. I also find that average complexity is lower in Mexico, and so is its employment of clerks. Finally, I find that the relationship between complexity and employment of clerks is stronger in Mexico, as it had been in the US in past decades when the price of IT was higher. On the other hand, the relationship between complexity and demand for managers, that has become increasingly important in the US, was still weaker in Mexico.

[3] Figure 3 also shows that the employment share of managers and other white-collar employees increased much more slowly over the same period.

Finally, I use the 1990 US data to examine a subgroup clerks engaged more directly in coordination of production. I find that the relationship between complexity and the employment share of these clerks is stronger than for other types of clerical workers.

Taken together, my results show that the complexity of the division of labor remains an important determinant of the employment of workers who process information. In 2000, an increase of standard deviation in complexity is associated with an increase of 1 percentage point (about 8 percent) in the employment of clerks, and 1.4 percentage points (about 14 percent) in the employment of managers.

The remainder of the paper is organized as follows. Section 2 presents a model that shows the effects of complexity on demand for information processing. Section 3 considers empirical evidence using US industries over time, two information technology revolutions, and a comparison of recent US and Mexican data. Section 4 concludes.

2 A Theory of Demand for Information Processing

To frame the key questions of this paper, I construct a framework for analyzing the effect of technology on the demand for information processing. I conceptualize the process of manufacturing as consisting of production and coordination. Production includes the tasks required to physically make the product. The workers who take part in production are operatives, craftsmen, laborers, and professional workers. Coordination involves generating, storing, and communicating information that is used to organize production. The information used to coordinate production is processed by clerks.[4]

All manufacturing firms need to perform the tasks required for production, but they can differ in their approach to coordination. One approach, which I call low information intensity, relies on informal information to coordinate production. For example, this may involve foremen supervising different sections of a plant. This approach typically involves costly misallocation of resources in production. The second approach relies on high information

[4] Managers' role can be seen as using the information processed by clerks to coordinate production. For simplicity, their tasks are not modeled in this framework.

intensity in coordination. This requires the firm to employ clerks, who use IT equipment to provide information that ensures a precise allocation of inputs in production.

Assume firms use labor inputs to produce output using a constant elasticity of substitution (CES) production function:

$$Y_{i,j} = \left[\int_0^{t_{i,j}} L_{i,j}(x)^{(\sigma-1)/\sigma} \, dx \right]^{\sigma/(\sigma-1)}, \tag{1}$$

where the information intensity i is either high (h) or low (l) and the industry is indexed by j. The complexity of the production function, $t_{i,j}$, is the measure of the set of different tasks employed in production. For simplicity, this model considers only labor inputs, rather than capital inputs, so complexity is closely related to the division of labor. In industry j, a set of workers of measure $L_{i,j}(x)$ is assigned to task x. The elasticity of substitution between the different tasks is σ, and I assume that tasks are gross substitutes ($\sigma > 1$).

I assume that firms using high information intensity execute all tasks correctly, so $t_{h,j} = t_j$. However, firms using low information intensity do not execute all tasks correctly. This may be due to problems in assigning workers to tasks, providing the right complementary inputs, or tracking inventories.[5] I assume that $t_{l,j} = t_j^{\gamma}$, where γ measures the efficiency of coordination in a low information intensity environment ($\gamma \in [0, 1)$). For example, if the content of tasks is easily comprehensible, it may be possible to coordinate work quite effectively even without much information, so γ may be close to 1.

A firm using high information intensity needs to gather information I_j in proportion to the size of its workforce and the complexity of the technology it uses:

$$I_j = \left(\int_0^{t_j} L_{h,j}(x) \, dx \right) t_j \tag{2}$$

The assumption that a more complex division of labor requires more information is a key part of the model. Consider two factories that employ an equal number of workers; in one factory

[5]For a related discussion see Crémer, Garicano, and Prat (2006).

all the workers perform the same task, while in the other each performs a different task. In order to effectively coordinate production in the more complex factory, labor has to be allocated across tasks, different complementary inputs need to be acquired and distributed, and various inventories need to be tracked. Thus, an effective coordination of a more complex labor force requires more information.

In order to analyze the demand for information processing, I assume that information is an intermediate input produced using CES technology:

$$I_j = \left[\alpha \left(L_{h,j}^c\right)^\eta + \beta \left(K_{h,j}^c\right)^\eta\right]^{1/\eta}, \tag{3}$$

where $L_{h,j}^c$ and $K_{h,j}^c$ are clerks and IT equipment. The productivity of clerks and IT equipment in producing information is captured by α and β, and the elasticity of substitution between clerks and information technology is $\sigma_I = 1/(1-\eta)$. Given a technology t_j, a firm producing with high information intensity solves the following optimization problem:

$$Min\left\{wL_{h,j} + w_c L_{h,j}^c + pK_{h,j}^c\right\} \quad s.t.: (1) - (3), \tag{4}$$

where the wage of labor is w, the wage of clerks is w_c and the price of information technology equipment is p. Using the simplifying assumption that all tasks in production are equally important, $L_{h,j}(x) = L_{h,j}/t_j$, and total labor demand is:

$$L_{h,j} = t_j^{-1/(\sigma-1)} Y_j \tag{5}$$

The demand for clerks and for information technology equipment ("clerical capital") is:

$$L_{h,j}^c = (\alpha + \beta\phi)^{-1/\eta} t_j^{(\sigma-2)/(\sigma-1)} Y_j \tag{6}$$

$$K_{h,j}^c = (\beta + \alpha/\phi)^{-1/\eta} t_j^{(\sigma-2)/(\sigma-1)} Y_j, \tag{7}$$

where $\phi \equiv (\beta w_c/\alpha p)^{\eta/(1-\eta)}$. Hence, the cost of production with high information intensity is:

$$C_{h,j} = t_j^{-1/(\sigma-1)} w Y_j + (\alpha+\beta\phi)^{-1/\eta}\, t_j^{(\sigma-2)/(\sigma-1)} w_c Y_j + \\ + (\beta+\alpha/\phi)^{-1/\eta}\, t_j^{(\sigma-2)/(\sigma-1)} p Y_j. \tag{8}$$

As long as the elasticity of substitution is sufficiently low ($\sigma \leq 2$), the cost function is decreasing in the level of complexity t_j, because the increased efficiency of production with a more complex technology outweighs the additional cost of information associated with it.[6]

A firm may choose to rely on low information intensity for supervising work. In this case it fails to execute some tasks correctly, but it does not need to employ clerks or IT equipment. The labor demand of a firm producing with low information intensity is:

$$L_{l,j} = Y t_j^{(\sigma-\gamma\sigma-1)/(\sigma-1)}, \tag{9}$$

and its cost function is:

$$C_{l,j} = t_j^{(\sigma-\gamma\sigma-1)/(\sigma-1)} w Y_j. \tag{10}$$

Cost decreases in t_j as long as the elasticity of substitution between tasks is sufficiently low ($\sigma < \frac{1}{1-\gamma}$). If $\sigma > \frac{1}{1-\gamma}$ then firms using low information intensity prefer the simplest technology available.

Using the constant returns to scale property of the production function we can define the net benefit of switching from low to high information intensity per unit of output as the difference between the marginal cost of a unit of output under both strategies:

$$I_{\Delta,j} \equiv MC_{l,j} - MC_{h,j} = (t_j^{(\sigma-\gamma\sigma-1)/(\sigma-1)} - t_j^{-1/(\sigma-1)})w - \\ - (\alpha+\beta\phi)^{-1/\eta}\, t_j^{(\sigma-2)/(\sigma-1)} w_c - (\beta+\alpha/\phi)^{-1/\eta}\, t_j^{(\sigma-2)/(\sigma-1)} p \tag{11}$$

[6] If $\sigma > 2$ then firms using high information intensity prefer a lower level of technology. I assume that this is not the case.

Firms employ clerks if and only if $I_{\Delta,j} > 0$. Only the firms using high information intensity employ clerks and information technology equipment, so a decline in the price of these inputs makes high information intensity more advantageous. In particular, if firms are using low information intensity, then a fall in the price of IT equipment makes high information intensity more profitable:

$$\frac{\partial I_{\Delta,j}}{\partial p} < 0. \tag{12}$$

Hence, if the price of IT equipment declines, firms may switch from low information intensity to high information intensity.

Now consider the case where all firms use high information intensity. We can compute the ratio of clerks to other workers:

$$c_j \equiv L_{h,j}^c / L_{h,j} = \left[\alpha + \beta^{1/(1-\eta)} \left(w_c/\alpha p\right)^{\eta/(1-\eta)}\right]^{-1/\eta} t_j. \tag{13}$$

When all firms use high information intensity we obtain the following comparative statics:

$$\frac{\partial c_j}{\partial t_j} > 0, \frac{\partial c_j}{\partial w_c} < 0, \frac{\partial c_j}{\partial p} > 0, \frac{\partial^2 c_j}{\partial p \partial t_j} > 0. \tag{14}$$

Similarly, we can define the intensity of IT equipment as:

$$k_j^c \equiv K_{h,j}^c / L_{h,j}. \tag{15}$$

so when information intensity is high:

$$\frac{\partial k_j^c}{\partial t_j} > 0, \frac{\partial k_j^c}{\partial w_c} > 0, \frac{\partial k_j^c}{\partial p} < 0, \frac{\partial^2 c_j}{\partial p \partial t_j} < 0. \tag{16}$$

We can interpret these results in the following way. First, if firms use high information intensity, more complex industries use more clerks and more IT equipment. Section 3.1 tests this prediction using variation across US manufacturing industries throughout the 20th century. This data also allows us to examine how the employment of clerical workers, relative to other workers, varies with the complexity of the division of labor.

Second, consider a sharp fall in the price of IT equipment; such a decline may cause firms to switch from low information intensity, where no clerks are used, to high information intensity. Once information intensity is high, more firms with more complex industries use more clerks. Thus, if complexity reflects persistent technological differences, industries that were more complex before the fall in the price of IT increase their employment of clerks more than other firms. Section 3.2 considers an early IT Revolution that took place from 1880-1910, and reduced the effective price of IT equipment. I examine how this change caused firms to increase their information intensity and their use of clerks, and how these changes were more pronounced in more complex industries.

Third, once firms use high information intensity, a further decline in the price of IT equipment causes them to substitute IT equipment for clerks; moreover, this substitution takes place differentially more in complex industries. In Section 3.3 examine such a change using the recent computer revolution. In absence of information on industry-level use of IT equipment, I use the employment of computer personnel as a proxy for the use of such equipment.

Finally, my interpretation of the model assumes that differences in complexity reflect differences in complexity. Section 3.4 compares the US industries with Mexican industries in 1990. I examine whether the measures of industry complexity are correlated across the two economies, and whether variations in complexity correspond to the fraction of clerks employed.

I begin my examination of complexity by defining the empirical measure of complexity and examining its relationship to the employment of clerks.

3 Empirical Evidence

3.1 US Industries Over Time

Technological complexity is a measure of the variety of tasks required to physically make products. In terms of labor inputs, complex production processes involve an occupationally diverse workforce. Accordingly, I define complexity as 1 minus the Herfindahl index of occupations of employees in each industry. This measure is designed to reflect occupational heterogeneity in production, but not in coordination; therefore, I calculate this measure excluding managers, clerks, accountants and auditors. My main source of data is the 1 percent Integrated Public Use Microdata Series (IPUMS) of the decennial censuses (Ruggles et al. 1997). Since occupational definitions change over time, I use the IPUMS 1950 classification of occupations and industries throughout the paper.[7] The Data Appendix discusses the construction of samples and the complexity measure.

Appendix Table A1 shows the industries with particularly high or low levels of complexity from 1880-2000. During the early 20th century complexity was positively correlated with using continuous-process or batch-production technologies (Chandler 1977; Goldin and Katz 1998), as industries using these technologies ranked above the median level of complexity.[8] In subsequent decades transportation industries ranked among the most complex industries, while textile and clothing industries were among the least complex.

Comparisons of industries and occupations over time should be taken with caution, even though they rely on a fixed classification of occupations. With this caveat in mind, note that complexity appears to have risen over time (Table 1). Complexity rankings are also very consistent over time: from 1880-1940 the decade-by-decade correlation of complexity ranged from 0.5-0.9, and since 1940 it has exceeded 0.9.

In order to examine the relationship between complexity and information processing, I

[7] Computer-related occupations are the only exception to this rule, since they are only defined from 1970 onwards.

[8] Industries using continuous-process or batch-production technologies include beverages, dairy products, grain-mill product, paints and varnishes and petroleum refining.

investigate the employment of clerks. An examination of clerical occupations shows that they use IT equipment to generate, store, and communicate information, which is used for coordination. Throughout the 20th century clerical occupations in manufacturing included stenographers, typists, secretaries, bookkeepers, shipping and receiving clerks, and office machine operators. In recent decades, clerks have often also been employed in record processing, material recording, scheduling and distribution, providing information, and message distribution (Hunt and Hunt 1986).

Clerks differ from other white collar workers, since their main task is to process information that is used to coordinate production. By contrast, engineers and other professional workers typically process information that is used to design products or a processes. Sales workers may also process information as part of their work, but their main objective is to sell products. Clerks also differ from managers, since managers perform actual coordination and supervision. However, during the 19th century the distinction between clerks and managers was not as clear as it is today, as managers often processed information themselves. Therefore, in the empirical analysis I consider the differential effect of complexity on managers and clerks.

Tasks of processing information are typically skill-intensive, so clerks are typically more skilled than most other manufacturing workers. The earliest nationally representative micro data that identifies the education and occupation of individuals is the 1940 Census. Appendix Table A2 shows the education of clerks and other workers, over age 50, who were employed in manufacturing in 1940. Clerks had much higher levels of education than other workers, suggesting that even if we account for changes in occupation and industry and attrition from the labor force, clerks were highly educated workers in the early decades of the 20th century.[9] Table A2 also shows that the skill gap between clerks and other workers persisted at least through 1980, although it has narrowed over time.

The model predicts that when firms use high information intensity, more complex indus-

[9] Goldin and Katz (1999) show that clerks were much better educated than blue collar workers in Iowa in 1915.

tries employ more clerks relative to other workers. In order to test this hypothesis I use individual-level census data to estimate a linear probability model:

$$c_{ij} = \alpha t_j + z_i' \beta + \varepsilon_{ij}, \tag{17}$$

where c_{ij} is an indicator for person i, working in industry j, being employed as a clerk (or a manager). The regressor of interest is industry j's complexity, t_j. I also include individual level geographic controls (z_i).[10]

The top panel of Table 2 shows results of cross-section regressions without industry-level controls for each decade from 1900-2000. These results suggest that since 1910, more complex industries have employed more clerks. An increase of one standard deviation in complexity was associated with an increase of about 1.6 percentage points in the employment of clerks in 1910. This partial correlation increased to about 2.9 percentage points in 1960 and declined to about 1 percentage point in 2000. In other words, from 1910-1980 a change in one standard deviation of complexity corresponded to 15-25 percent of the average clerical employment in manufacturing, declining to 8-13 percent in 1990 and 2000. The results are similar in magnitude, though less precise, after controlling for the capital/labor ratio. Table 2 also shows that managerial employment, unlike clerical employment, was not correlated with higher complexity in 1910 and 1920, and it was substantially weaker throughout the first half of the 20th century.[11] This lends evidence that clerks and managers performed different tasks even in the early 20th century, so the demand for their services was driven by different factors.

The evidence presented thus far is consistent with the hypothesis that the coordination of a more complex production process requires relatively more clerks, but I cannot rule out

[10] Specifications controlling for industry-level ratios of capital to labor give similar coefficients for 1910 and 1960-1990. However, the estimate for 1910 is not significant at the 5 percent level.

[11] The relationship between complexity and managerial employment became significant from 1940 onwards, with a change of one standard deviation in complexity associated with 0.4-1.5 percentage points (10-20 percent change) in managerial employment. Interestingly, the correlation between complexity and managerial employment has strengthened in recent decades.

reverse causality. Since the ranking of industries by complexity is stable over time, it is difficult to find an exogenous source of variation in complexity. I therefore examine if more complex industries increased their demand for clerks more rapidly as an early decline in the effective price of IT equipment caused firms to switch from low information intensity to high information intensity.

3.2 An Early IT Revolution: 1880-1910

This section discusses the early IT revolution, which took place from 1880-1910. I show that this revolution increased the demand for clerks, and that industries that were more complex to begin with increased their demand more rapidly. I also show that the demand for clerks cannot be explained by variations in plant size across industries.

Yates (1989) and Levenstein (1991) describe the IT equipment that came into widespread use in offices from 1880-1910. The new IT equipment had substantial impact on four different aspects of information processing: production, copying, storage, and communication.[12]

First, the production of information was enhanced by the advent of the typewriter, which allowed rapid and inexpensive generation of neat documents. The patent that led to the first commercially viable typewriter was registered in 1868. During the 1870s the first typewriters were sold to court reporters and telegraphers. Once the value of typewriters was realized by firms, their production expanded rapidly, and in 1900 alone over 140,000 machines were sold. The average annual growth rate of the real value of typewriters produced in the U.S. from 1890-1909 was 9 percent, and it fell to 3 percent over the subsequent five years (see Figure 2). Complementary innovations, such as the visible typewriter and the dictaphone, allowed typists to see their output as they typed and to record dictations. Cash registers and other office machines also became popular, and from 1890-1909 the real value of such equipment grew at an average rate of 16 percent; the growth rate fell to a mere 3 percent over the subsequent five years.[13]

[12] Broadberry and Ghosal (2002) consider the productivity consequences of the adoption of these office technologies.

[13] The data on output comes from Roy (1990). Following Goldin and Katz (1995), Table 5, I use the

Second, the technology for copying information also improved from 1880-1910. Earlier copying technologies, such as the book press and the rolling press, suffered from substantial drawbacks, and carbon paper was incompatible with contemporary pens. With the advent of typewriters, carbon paper allowed clerks to produce several copies at the point of origin. This innovation reduced copying cost and made documents neater and more durable. Stencils, first patented in 1876, became useful for mass duplication of internal communication. The first photocopying machine was invented in 1900, and in 1911 the Taft Commission evaluated it and found it useful for repeated copying of complicated documents and diagrams. The widespread use of technologies for producing and copying information increased the use of paper. From 1890-1910, the tonnage of fine writing paper manufactured in the U.S. grew at a rate of 6.4 percent per capita, compared to 4.8 percent from 1910-1930 (Feenberg and Miron 1997).

Third, the technology for storage and retrieval of information was quite limited before the last decades of the 19th century. Documents were typically stored in pigeonhole desks that could only accommodate few documents, or in cabinets and flat files that made retrieval difficult. Vertical filing was presented to the business world at Chicago's World Fair of 1893. It allowed systematic grouping of documents, quick rearrangement and retrieval, and substantial saving in storage space.

Finally, the technology for communicating information also improved from 1880-1910. The telegraph was introduced to America as early as the 1840s, but it proved less useful in manufacturing than in the railroad industry (Chandler 1977). More important for manufacturing was the telephone, which was invented in 1876 and came into commercial use shortly thereafter. By the 1890s private branch exchanges were used in cities across the U.S., and around 1900 long distance communication become possible. Figure 2 shows that the annual growth rate in phones per capita from 1880-1910 was about 16 percent; in the next decade it fell to about 4 percent.[14] According to data from the 1920s to the 1970s, about one-third of

cost-of-living index due to Rees as the price deflator (Historical Statistics 1997, series E186).

[14]Similarly, Sorkin (1980) finds that the number of telephone connections made by the Bell Company

the phones in the U.S. were used by businesses, and this fraction declined only slowly over time (Historical Statistics 1997). The use of printed communication also grew rapidly: the annual growth rate of the number of items sent by the U.S. Postal Service was about 5.4 percent from 1874-1910 and only about 1.6 percent from 1910-1970.[15]

Taken together, this evidence shows that an early information technology revolution took place from 1880-1910. The key period for this revolution was the first decade of the 20th century, when technologies for producing, copying, storing and communicating information matured. This early IT revolution created the potential to use information much more intensively than before.

Throughout most of the 19th century manufacturing firms made little systematic use of information for coordination of production. Rather, firms were often managed by owners, typically with the help of skilled artisans or foremen (Chandler 1977; Hounshell 1984; Yates 1989; Nelson 1995). To be sure, there were exceptional firms that employed innovative techniques of cost accounting; but these techniques were typically firm-specific and suffered from substantial shortcomings (Fleischmen and Tyson 1993).

During the late 19th century, a group of managers and engineers led a movement for "Systematic Management" (Nelson 1980, 1995).[16] They promoted systematic cost accounting, use of job cards and time clocks, inventory control, centralized purchasing, and incentive wages. Management literature had barely existed before 1870, but it expanded rapidly during the next thirty years. It was also around the turn of the century that several American universities first established programs in business administration. The next step in expanding the use of information was taken by the movement for "Scientific Management", led by Frederick W. Taylor. Taylor and his associates served as consultants and initiated changes in plant management practices. These changes included mechanical innovations, improved cost

increased from 1.87 billion to 8.14 billion from 1900-1910. In per capita terms, this reflects an average annual growth rate of about 14 percent during the first decade of the 20th century, compared to 2.3% in the subsequent decade and an average growth rate of about 3.5% from 1920-1970.

[15] Author's calculations are based on data from Bangs (1875) and Sorkin (1980).

[16] The availability of new power sources (steam and later electric power) also allowed more flexibility in organizing the factory floor, and complemented these changes.

accounting and purchasing systems, the introduction of storerooms and planning departments, use of time studies, and more sophisticated incentive wage systems. While neither universally adopted nor entirely successful, these changes were a sign of the attempts to use information for managerial purposes.

Most of the literature on this organizational transformation has focused on the managers who initiated them (e.g. Chandler 1977 and Nelson 1995), but the role of clerks in facilitating these changes was no less important. As Table 1 shows, production workers accounted for about 95 percent of the manufacturing workforce in 1880, while most of the remaining 5 percent were managers and owners. But by 1910 clerks emerged as a new class of white collar employees in the manufacturing industries, outnumbering managers by about two-to-three.[17]

Existing data on wages suggest that on average clerks earned almost twice as much as other manufacturing workers in 1890, reflecting their higher level of education at a time when high school graduates were scarce.[18] As Figure 4 shows, the wage of clerks relative to other manufacturing workers appears to have remained constant or even increased slightly from 1890-1910, while clerical employment increased rapidly. This consistent with the view that the changes in clerical employment reflect a large increase in the relative demand for clerks.

The growth in demand for clerks raised the demand for skilled workers when high-school education was still scarce (Goldin and Katz 1995), and this change may have contributed to the onset of the high-school movement in subsequent decades (see Goldin 1998). The increased demand for clerks also had important implications for the employment of women: although men filled the majority of new clerical jobs in manufacturing created by 1910,

[17] Estimates using the IPUMS data and the 1950 occupational definitions suggest that there were about 18,000-20,000 clerks in manufacturing in 1870 and 1880. This figure corresponds to about 1 percent of the manufacturing labor force. The 1950 classification of clerical occupations appears to match earlier classification systems well only for data from 1880 onwards, so estimates prior to 1880 are potentially subject to substantial measurement error. The Census of Manufactures uses a narrower definition of clerical occupations than the population census. According to these definitions there were fewer than 6,000 and 12,000 clerks in manufacturing in 1870 and 1880.

[18] See Appendix Table A2 calculations based on Goldin and Katz (1998).

clerical work became feminized after 1910.

In order to examine how the changes in demand for clerks varied by industry, I use the measure of complexity introduced in the previous section. The model predicts that more complex industries increase their employment of clerks more rapidly than other industries, assuming that complexity ranking of industries is stable over time. To test this hypothesis I estimate a pooled cross-section regression for 1880, 1900, and 1910:

$$c_{ijt} = \alpha(t_{j,1880} * I_t) + x_j'\beta + z_i'\gamma + d_j'\delta + T_t'\lambda + \varepsilon_{ijt}, \tag{18}$$

where $t_{j,1880}$ is the complexity of industry j in 1880, I_t is a dummy for 1910 (after the information technology revolution had matured), x_j includes industry-level controls, x_j controls for industry level covariates, d_j is a vector of industry fixed effects, and T_t is a vector of year dummies.

As the results in Table 3 show, industries that were more complex in 1880 increased their employment of clerks more rapidly than other industries over the next three decades. These results are robust to different weighting methods (columns 1 and 2). Column 3 limits the sample to industries with large plants. The coefficient of interest, α, is larger and more precise, suggesting that my results are not driven by industries with small and heterogenous plants. In a similar vein, I exclude industries that are loosely defined (e.g. miscellaneous wood products), and the results are similar to the benchmark. The results are also robust to including only industries with more than 100 observations in 1880 (column 5). Finally I use an alternative measure of complexity, computed using only production workers; once again, the results are similar (column 6). These results suggest that an increase of one standard deviation of complexity in 1880 corresponds to an increase of about 1-2 percentage points (or about about 15-30 percent) in the employment of clerks in 1910, similar to the cross-sectional estimates.[19] By comparison, the estimated effect of complexity on the relative employment

[19] Cross-section and panel regressions using industry variations in average plant size show that it is uncorrelated with employment of clerks. However, I cannot rule out the possibility that the growth in firm size discussed in Chandler (1977) may have contributed to the increased demand for clerks.

of managers is positive but substantially smaller, and only marginally significant.

This evidence suggests that complexity, rather than plant size, was the important determinant of demand for clerks around the turn of the 20th century. The next section examines how complex industries responded to changes in the effective price of IT equipment that took place in recent decades.

3.3 A Recent IT Revolution: 1960-2000

The computer revolution that has taken place in recent decades shares much in common with the IT revolution that occurred around the turn of the 20th century. First, computer software (e.g. word processors and electronic spreadsheets) and hardware (e.g. printers) substantially reduce the cost of producing information, just as the typewriter had done almost one century earlier. Second, computers allow us to copy information at almost no cost. Third, computers allow effective storage and retrieval of vast amounts of information. And finally, computer networks make communication cheaper and more effective, as the telephone had done in the past.

In 1970 there were fewer than 75,000 computers in the U.S., but over the next decade about 875,000 computers were purchased (Phister 1979; Hunt and Hunt 1986). Substantial reduction in the cost of processing power (Jorgenson 2001; Nordhaus 2001) induced a transition from mainframe computers to microcomputers. The range of computer applications also increased over time. By 1984 nearly half of the clerks in the U.S. used computers, compared to only a quarter of other workers; more recent data show that the fraction of clerks (and other workers) using computers has since continued to rise (Friedberg 2003). I therefore conclude that during the 1970s (and possibly the early 1980s) the price of office computers declined sharply, and we can examine the effect of this decline on the demand for clerks.

The new IT equipment reduced the demand for clerks in three different ways (Hunt and Hunt 1986; Osterman 1986). First, new equipment has directly replaced clerical workers . For example, automatic switching replaced telephone operators; office dictation equipment took the place of stenographers; computer software for pricing insurance supplanted specialized clerks (raters); automatic mail sorting devices replaced mail clerks; and computerized inventory management software took the place of shipping and receiving clerks. Second, indirect replacement occurred as the new equipment raised the productivity of clerical work, reducing the number of clerks required to perform a given task. Finally, reorganization of information processing has also reduced the demand for clerks. For example, some stenographers and typists lost their jobs when word processing allowed workers to produce information cheaply and efficiently.[20]

As Table 1 shows, the fraction of clerks in the manufacturing workforce fell from 14.4 percent in 1960 to 11.4 percent in 2000. This decline does not reflect an increase in the relative wages of clerks. As Appendix Table A3 shows, the wage of clerks relative to other manufacturing workers has gradually fallen over this period. Taken together, this evidence suggests manufacturers' demand for clerk, relative to other workers, has been falling.

Of particular interest is the effect is the differential response of complex industries to the fall in the price of IT equipment. The first four columns in Table 4 show the cross-sectional relation between complexity and the employment of computer personnel from 1970-2000.[21] The coefficient is positive in all years, although it is statistically significant at the 5 percent level only for 1980 and 2000 and marginally significant for 1970 and 1990. These results are consistent with the model's prediction that more complex industries have higher demand for information technology equipment.

Given the persistence of complexity, the model predicts that more complex industries substitute IT equipment for clerks more rapidly than other industries. Assuming that the number of computer operators required for each computer does not vary across industries in each year, we expect that higher complexity leads to greater employment of computer

[20] Osterman (1986) argues that these factors might be at least partially offset by "information deepening", as firms may use more information relative to other intermediate inputs.

[21] See Data Appendix for a definition of the category of "computer personnel."

personnel. Table 4 shows the results of the following regression using data from 1960-1990:

$$c_{ijt} = \alpha(t_{j,1960} * I_t) + x_j'\beta + z_i'\gamma + d_j'\delta + T_t'\lambda + \varepsilon_{ijt}, \tag{19}$$

where $t_{j,1960}$ is the complexity of industry j in 1960, I_t is a dummy for 1980 and 1990 (after the computer revolution had matured), x_j includes industry-level controls, d_j is a vector of industry fixed effects and T_t is a vector of year effects. The results suggest that more complex industries significantly reduced their employment of clerks. An increase of one s.d. in complexity in 1960 is associated with a decrease of about 1.3 percentage points in clerical employment in 1980 and 1990.

Table 4 also shows that complex industries increased the demand for computer personnel more than other industries. However, this differential increase was only about 1/20 in magnitude compared to the decline in their employment of clerks. In part, this finding may reflect the fact that a single computer operator may be in charge of multiple computers or multiple user interfaces. It may also reflect the capacity of a single computer operator to substitute multiple clerks that are not using computers. Finally, it is also possible that some jobs may have been outsourced to service providers outside manufacturing.[22]

In summary, these findings show that manufacturing industries substituted computer personnel for clerks, and that this substitution was significantly stronger among more complex industries. The next section compares the relationship between complexity and demand for information processing in the US and Mexico.

3.4 Comparing USA and Mexico in 1990

In this section I use data from IPUMS International to examine the relationship between complexity and demand for information processing in the US and Mexico in 1990. This approach sheds light on the hypothesis that differences in complexity are largely due to

[22] Interestingly, complex industries also appear to have increased their demand for managers more than other industries. This finding suggests that they may face more challenges in coordinating complex production processes, even when plenty of information is available.

underlying technological differences, and it allows us to compare the differences in complexity and information processing across economies that are in different stages of development. I also use the 1990 US data to examine a subgroup of clerks engaged more directly in coordination, and I test whether it is

The top panel in Table 5 compares manufacturing industries in the US and Mexico in 1990. The data uses harmonized definitions of industries, but each country's occupational definitions are different (See Data Appendix).[23] Despite these differences, the summary statistics are consistent with the hypothesis that differences in complexity reflect technological variations. The average complexity in Mexico is lower than the US; industry-by-industry comparisons show that the complexity of Mexican industries is higher in only 8 of the 44 industries. Moreover, the correlation in industry-level complexity across the two countries is about 0.3.

Consistent with the hypothesis that complexity affects demand for information processing, The Mexico had fewer clerks than the US and also fewer managers. As the bottom panel of Table 5 shows, the relationship between complexity and the fraction of clerks (and managers) was highly significant in both economies. Interestingly, in Mexico, complexity was more strongly associated with clerical employment, as it had been in the US in past decades. By contrast, in the US complexity was more strongly associated with managerial employment. Although a full investigation of these issues is outside the scope of the current paper, this finding suggests that in the highly complex production environment we face today, the cost of a more detailed division of labor may be increasingly due to the difficulty of analyzing information and coordinating production, rather than processing information.

Like all the analysis thus far, the comparison of US and Mexican data relies on a broad definition of clerical occupations. Although the tasks of most of the clerical occupations are defined in broad terms (e.g. secretaries, general office clerks), the occupational classifications

[23] The total number of occupations in the two countries is very similar. When excluding managers, clerks, and misclassified workers there are 354 occupations in the US data and 367 occupations in the Mexican data. Note also that the results are robust to restricting the US sample size to the same size as the Mexican data.

of the 1990 US data identify some clerical occupations ("coordination" clerks) that are particularly related to processing information that is used to coordinate production. We can compare those "other" clerical occupations that are less related to the coordination of production, and to the larger residual category of "general" clerks.[24]

As Table 6 shows, complexity is most strongly associated with the employment of coordination clerks. One standard deviation in complexity is associated with an increase of 11 percent in the employment share of coordination clerks, compared to about 6 percent and 4 percent of general and other clerks. This evidence provides further support for the relationship between coordination and demand for information processing.

4 Conclusions

In this paper I analyze the cost of processing information that is used to coordinate complex production processes. One contribution of this paper is in constructing and analyzing a measure of the complexity of the division of labor, which measures the extent of occupational heterogeneity in an industry's workforce. Using US data on manufacturing industries over more than a century I find evidence that this measure is highly persistent, likely reflecting underlying differences in the production process across industries. A comparison of recent data from US and Mexico reveals that US industries are more complex, and within each country the rankings of industries by complexity are similar.

I find evidence that more complex industries employ relatively more clerks, and that this correlation is consistent and robust for almost a century. Mexican data for 1990 shows similarly strong evidence.

In order to further document the effect of complexity on the demand for clerks, I examine an early IT Revolution that took place from 1880-1910. The sharp decline in the effective price of IT induced firms to use information more intensively, by employing both clerks and

[24]"Coordination" clerks are: production coordinators; stock and inventory clerks; personnel clerks; and payroll and timekeeping clerks. "Other" clerks are: transportation ticket and reservation agents; receptionists; order clerks; billing clerks; dispatchers; weighers, measurers, and checkers; and investigators and adjusters, except insurance. The residual catgory includes the rest of the clerical occupations.

IT equipment. This finding suggests that by increasing the demand for clerks, technological change may have contributed to the onset of the High School Movement from 1910 onwards and to women's integration into the labor force. Interestingly, I find that industries which were more complex in 1880, when there was very little employment of clerks, increased their demand for clerks significantly more than other industries.

Whereas in the past clerks and IT equipment were complements, in recent decades they have become substitutes. I find that manufacturing firms have substituted computers for clerks, reducing the employment share of clerks despite the decline in their relative wages. This result provides a concrete case in which technology and a specific set of skills are substitutes rather than complements. Moreover, my findings suggest that the replacement of clerks by computers was more rapid in more complex industries.

Lastly, this paper assesses the extent to which complexity increases demand for clerks. I find that one standard deviation in complexity is associated with an increase of 1 percentage point (about 8 percent) in the employment of clerks. In recent decades, complexity has also become strongly associated with the employment of managers: an increase of 1 s.d. in complexity is associated with an increase of 1.4 percentage points (about 14 percent) in the employment of managers. The results suggest that the complexity of the production process is an important determinant of the demand for white-collar workers.

Data Appendix

The data for this paper comes primarily from the Integrated Public Use Microdata Series (IPUMS) samples of the U.S. decennial household census. This data includes samples of approximately 1 percent of the decennial US census from 1860-2000, excluding 1890 and 1930, for which no data is currently available. This data allows a systematic cross-sectional and longitudinal analysis of the manufacturing workforce, while controlling for geographic variation. Using the IPUMS 1950 classification of occupations and industries mitigates some of the problems that could arise from changes in the classification of jobs throughout the period.[25] Nevertheless, before 1880 people were not directly asked for both their occupation and industry, so the distinction between the two had to be imputed and this created a potential for substantial measurement error. This data also lacks the wage and educational attainment of workers before 1940.[26] In order to obtain wage data for that period, I rely on the estimates of Goldin and Katz (1995), which are based on partial data.

I include workers in all manufacturing industries, except for 'miscellaneous manufacturing industries' and 'not specified manufacturing industries'. Dropping these observations reduced the sample by about 5 percent. Specific occupations that appeared in the data but seem incompatible with employment in manufacturing were excluded from the sample. These occupations include: chiropractors; clergymen; dancers and dancing teachers; entertainers; farm and home management advisors; foresters and conservationists; funeral directors and embalmers; recreation and group workers; religious workers; therapists and healers; farmers (owners and tenants); farm managers; auctioneers; hucksters and peddlers; real estate

[25] In the classification of industries, I merged together several industries, in order to facilitate the merging of the household census data with the census of manufactures data. Specifically, I merged industries 306 (logging) and 307 (sawmills, planing mills, and mill work); industries 336 (blast furnaces, steel works, and rolling mills) and 337 (other primary iron and steel industries); industries 346 (fabricated steel products) and 347 (fabricated nonferrous metal products); industries 439 (yarn, thread, and fabric mills), 436 (miscellaneous textile mill products) and 449 (miscellaneous fabricated textile products); and industries 399 (miscellaneous manufacturing industries) and 499 (not specified manufacturing industries). I also dropped all the observations that are classified as: employed, unclassifiable; non-occupational response; occupation missing/unknown; and N/A (blank).

[26] The schooling variable does make it possible to estimate high school participation across geographical regions.

agents and brokers; stock and bond salesmen; all service workers; all farm laborers; people whose occupation was unclassified; people with a non-occupational response; and all people with occupation missing/unknown or non-available. Dropping all these occupations further reduced the sample by about 3 percent.

The only industry-level variable constructed using the IPUMS data is the complexity measure. It is defined as one minus the Herfindahl index of occupations, excluding managers, clerks and accountants and auditors. In some specifications I also use an alternative measure of complexity, calculated using only production workers (Craftsmen, Operatives, and Laborers not elsewhere classified).

I construct other industry-level variables using the Census of Manufactures data for the years 1880-1909 and 1960-1990. The data for the early period includes the number of workers and establishments, as well as the value of output and cost of various inputs for over 250 industries, as classified by Roy (1990). The more recent data is from the NBER-CES Manufacturing Industry Database (Bartelsman et al. 2000).

The household census and the Census of Manufactures data use different classifications of industries and the latter classification is typically much more detailed. In order to match the data I pooled together several of the IPUMS industries: logging was merged with sawmills, planing mills, and mill work; blast furnaces, steel works, and rolling mills was merged with other primary iron and steel industries; yarn, thread, and fabric mills was merged with miscellaneous textile mill products and miscellaneous fabricated textile products. Merging these industries yielded a total of 52 industries, although a few had no observations in some census years.

Under the IPUMS classification of occupations, all white collar workers are classified into one of four one-digit categories: clerks, managers, professional workers and sales workers. I constructed the category of computer personnel based on the IPUMS occupations for 1970-2000 since no such category existed prior to 1970. In 1970 this category includes computer and peripheral equipment operators. In 1980 and 1990 it includes supervisors, computer equipment operators; computer operators; and peripheral equipment operators. In 2000 it includes computer and information systems managers; database administrators; network and computer systems administrators; and computer operators.

The analysis in section 6 uses US and Mexican data on occupations and industries in 1990 from the Integrated Public Use Microdata Series-International. I exclude various US miscellaneous industries: Toys, amusement, and sporting goods; Miscellaneous manufacturing industries; and Manufacturing industries, n.s., and similar Mexican industries: Not sufficiently specified industries, and Other manufacturing industries. Appendix Table A4 shows the matching of US and Mexican industries. The calculations exclude managers, administrative workers, and farm workers.

References

Acemoglu, Daron. “Directed Technical Change”, Review of Economic Studies, 2002, p. 782-809.

Autor, David H., Frank Levy and Richard J. Murnane. "The Skill Content of Recent Technological Change: An Empirical Exploration", Forthcoming Quarterly Journal of Economics, 118(4), November 2003.

Bangs, George S. "A Letter to the Postmaster General showing the Cost, Revenues and Comparative Profit and Loss of Carrying the Different Classes of Mail Matter", Washington Government Printing Office, 1875.

Bartelsman, Eric J., Randy A. Becker, and Wayne B. Gray. "NBER-CES Manufacturing Industry Database", National Bureau of Economic Research, June 2000 http://www.nber.org/nberces/nbprod96.htm

Becker, Gary S. And Kevin M. Murphy. "The Division of Labor, Coordination Costs, and Knowledge", *Quarterly Journal of Economics* v107, n4 (November 1992): 1137-60.

Broadberry, Stephen and Sayantan Ghosal. "From the Counting House to the Modern Office: Explaining Anglo-American Productivity Differences in Services, 1870–1990." The Journal of Economic History, Volume 62, Issue 04, p. 967-998, Dec. 2002.

Carter, Susan B. et al., ed. H*istorical statistics of the United States on CD-ROM: colonial times to 1970*, U.S. Bureau of the Census.

Chandler, Alfred D., Jr. *The Visible Hand: The Managerial Revolution in the American Business*, The Belknap Press of the Harvard University Press, 1977.

Crémer, Jacques, Luis Garicano, and Andrea Prat. "Language and the Theory of the Firm." forthcoming, *Quarterly Journal of Economics*, 2006.

Davis, Margery W. *Women’s Place Is at the Typewriter – Office Work and Office Workers 1870-1930*, Temple University Press, 1982.

Duranton, Gilles and Hubert Jayet. "Is the Division of Labour Limited by the Extent of the Market? Evidence from French Cities." Mimeo, March 2006.

Feenberg, Daniel and Jeff Miron. "Improving the Accessibility of the NBER's Historical Data", Journal of Business and Economic Statistics, Volume 15 Number 3 (July 1997) pages 293-299. Data series: m01104a U.S. Fine Writing Paper Production 01/1918-12/1933

Fine, Lisa M. *The Souls of the Skyscraper – Female Clerical Workers in Chicago, 1870-1930*, Temple University Press, Philadelphia, 1990.

Fleischman, Richard K. and Thomas N. Tyson, "Cost Accounting during the Industrial Revolution: The Present State of Historical Knowledge", The Economic History Review, New Series, Volume 46, Issue 3 (Aug. 1993), p. 503-517.

Friedberg, Leora. "The Impact of Technological Change on Older Workers: Evidence from Data on Computer Use", *Industrial and Labor Relations Review*, Vol. 56, No. 3 (April 2003).

Garicano, Luis. "Hierarchies and the Organization of Knowledge in Production", *Journal of Political Economy* v108, n5 (October 2000): 874-904.

Garicano, Luis and Esteban Rossi-Hansberg. "Organization and Inequality in a Knowledge Economy." forthcoming, *Quarterly Journal of Economics*, 2006.

Garicano, Luis and Thomas Hubbard, "Specialization, Firms and Markets: The division of Labour Between and Within Law Firms," Centre for Economic Policy Research Discussion Papers, CEPR Discussion Papers 3699, 2003.

Goldin, Claudia, "America's Graduation from High School: The Evolution and Spread of Secondary Schooling in the Twentieth Century", *Journal of Economic History*, Vol. 58 (June 1998): 345-374

Goldin, Claudia and Lawrence F. Katz. "Education and Income in the Early 20th Century: Evidence from the Prairies." National Bureau of Economic Research, Inc, NBER Working Papers: 7217, 1999

Goldin, Claudia and Lawrence F. Katz. "The Decline of Non-Competing Groups: Changes in the Premium to Education, 1890 to 1940.", National Bureau of Economic Research Working Paper: 5202, 1995.

Goldin, Claudia and Lawrence F. Katz. ”The Origins of Technology-Skill Complementarity”, Quarterly Journal of Economics, 1998.

Hartman Strom, Sharon. "‘Light Manufacturing’: The Feminization of American Oddice Work, 1900-1930”, *Journal of Industrial and Labor Relations Review*, Volume 49, Issue 1, 1989, p. 53-71.

Hartman Strom, Sharon. *Beyond the Typewriter: Gender, Class and the Origins of Modern American Office Work, 1900-1930*, University of Illinois Press, 1992.

Hounshell, David A. *From the American System to Mass Production, 1800-1932*, Johns Hopkins University Press, 1984.

Hunt, H. Allan and Timothy L. Hunt. *Clerical employment and technological change*, W.E. Upjohn Institute for Employment Research, 1986.

Jorgenson, Dale W. "Information Technology and the U.S. Economy", *American Economic Review*, March 2001, p. 1-32.

Levenstein, Margaret. "The Use of Cost Measures: The Dow Chemical Company, 1890-1914" in *Inside the Business Enterprise: The Use and Transformation of Information*, ed. Peter Temin (University of Chicago Press), 1991.

Minnesota Population Center. Integrated Public Use Microdata Series-International: Version 2.0. Minneapolis: University of Minnesota, 2006.
www.ipums.org/international

Nelson, Daniel. *Fredrick W. Taylor and the Rise of Scientific Management*, The University of Wisconsin Press, 1980.

Nelson, Daniel. *Managers and Workers – origins of the Twentieth-Century Factory System in the United States, 1880-1920*, The University of Wisconsin Press, second edition (1995).

Nordhaus, William. "The Progress of Computing", Cowles Foundation Discussion Paper No. 1324, 2001.

Osterman, Paul. "The Impact of Computers on the Employment of Clerks and Managers", *Industrial and Labor Relations Review*, Volume 39, Issue 2, Jan. 1986, p. 175-186.

Phister, Montgomery, Jr. *Data Processing Technology and Economics*, Second Edition, Digital Press, 1979. [Table II.1.28]

Radner, Roy. "Hierarchy: The economics of managing." *Journal of Economic Literature*; Sep92, Vol. 30 Issue 3, p. 1382-1415.

Radner, Roy. "The Organiztion of Decentralized Information Processing." *Econometrica*, Sep93, Vol. 61 Issue 5, p1109-1146

Rotella, Elyce J. Women's Labor Force Participation and the Growth of Clerical Employment in the United States, 1870-1930 (in Summaries of Doctoral Dissertations), *The Journal of Economic History*, Vol. 39, No. 1, The Tasks of Economic History. (Mar., 1979), pp. 331-333.

Roy, William G. "Rise of American Industrial Corporations, 1880-1914" [Computer file]. Los Angeles, CA: University of California, Dept. Of Sociology [producer], 1990. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1990.

Ruggles, Steven et. al. *Integrated Public Use Microdata Series: Version 2.0*, Minneapolis: Historical Census Projects, University of Minnesota, 1997 (www.ipums.org).

Smith, Adam. *The Wealth of Nations*, 1776 (Reprinted in New York: Modern Library, 1965).

Sorkin, Alan L. *The Economics of the Postal System*, Lexington Books, 1980.

U.S. Department of Commerce, Bureau of the Census. *Historical Statistics of the United States, Bicentennial edition*, Cambridge University Press, 1997 (CD).

U.S. Census Bureau, Population Division, Education and Social Stratification Branch. "Computer Use in the Unitd States", October 1989 (P23-171) http://www.census.gov/population/www/socdemo/computer/p23-171.html.

Yates, JoAnne. "Control through Communication – The Rise of System in American Management", Johns Hopkins University Press, 1989.

Table 1. Descriptive Statistics

Year	Total manufacturing workforce (millions)	Fraction of white collar occupations in manufacturing						Fraction of women among clerks	Average industry complexity
		Clerks	Managers	Professional workers	Sales workers	Computer administrators and operators (see note)	All white collar workers		
1860	0.9	0.001	0.022	0.004	0.000	0.000	0.028	0.000	0.400
1870	1.6	0.012	0.031	0.004	0.000	0.000	0.047	0.089	0.438
1880	1.8	0.011	0.033	0.007	0.001	0.000	0.052	0.046	0.361
1900	3.8	0.026	0.035	0.010	0.001	0.000	0.071	0.173	0.583
1910	7.6	0.064	0.043	0.011	0.010	0.000	0.128	0.357	0.684
1920	9.1	0.083	0.039	0.015	0.002	0.000	0.139	0.450	0.688
1940	11.1	0.103	0.041	0.030	0.031	0.000	0.205	0.428	0.679
1950	14.6	0.113	0.047	0.042	0.023	0.000	0.223	0.542	0.714
1960	21.4	0.144	0.044	0.062	0.026	0.000	0.275	0.690	0.696
1970	24.4	0.140	0.045	0.081	0.023	0.002	0.290	0.697	0.716
1980	25.5	0.131	0.066	0.090	0.022	0.005	0.309	0.691	0.755
1990	23.4	0.124	0.096	0.121	0.030	0.007	0.370	0.696	0.791
2000	23.1	0.114	0.099	0.138	0.030	0.007	0.381	0.667	0.775

NOTES: The sample includes workers in the manufacturing industries, as described in the Data Appendix. Using the IPUMS 1950 classification of occupations, all white collar workers fall into one of the following categories: clerks, managers, professional workers and sales workers. The category of computer administrators and operators includes some clerks, managers and professional workers (see details in Data Appendix). The complexity measure is computed as one minus the Herfindahl index of occupations, excluding managers, clerks and accountants and auditors. Means are calculated using census person weights.

Table 2. Complexity and Industry-level Employment Shares of Clerks and Managers: 1900-2000

	1900	1910	1920	1940	1950	1960	1970	1980	1990	2000
					A. Clerks					
Complexity	-0.003	0.102	0.086	0.111	0.107	0.136	0.151	0.157	0.139	0.106
	-(0.020)	(0.041)	(0.029)	(0.029)	(0.029)	(0.031)	(0.033)	(0.045)	(0.049)	(0.044)
					B. Managers					
Complexity	-0.017	-0.010	-0.034	0.039	0.024	0.039	0.052	0.083	0.128	0.156
	(0.028)	(0.025)	(0.018)	(0.010)	(0.012)	(0.009)	(0.010)	(0.018)	(0.022)	(0.023)
Observations	19,080	30,018	90,504	112,424	170,925	214,951	243,983	255,021	237,953	233,649

NOTES: The sample includes workers in the manufacturing industries, as described in the Data Appendix. The complexity measure is computed as one minus the herfindahl index of occupations, excluding managers, clerks and accountants and auditors. All the regressions include state dummies. Dummies for the size of place in which the person lived were included in the years for which this information was available (1900, 1910, 1940, 1950, 1980, 1990). All regressions are weighted (using person-weights). Robust standard errors in parentheses are clustered at the industry level.

Table 3. Effect of Complexity in 1880 on Employment of Clerks and Managers in 1910

	(1)	(2)	(3)	(4)	(5)	(6)
			A. Clerks			
(Complexity in 1880) x (Year=1910)	0.048	0.051	0.082	0.037	0.049	0.066
	(0.019)	(0.019)	(0.015)	(0.019)	(0.020)	(0.017)
Ln(capital / labor)	-0.012	-0.014	-0.047	-0.010	-0.012	-0.019
	(0.012)	(0.015)	(0.018)	(0.012)	(0.012)	(0.012)
Ln(average establishment size)	-0.008	-0.005	-0.024	-0.010	-0.009	-0.009
	(0.006)	(0.007)	(0.005)	(0.006)	(0.006)	(0.006)
Year=1900	0.019	0.018	0.039	0.021	0.019	0.023
	(0.008)	(0.011)	(0.009)	(0.008)	(0.008)	(0.008)
Year=1910	0.034	0.032	0.062	0.039	0.033	0.036
	(0.012)	(0.015)	(0.013)	(0.013)	(0.012)	(0.014)
			B. Managers			
(Complexity in 1880) x (Year=1910)	0.017	0.018	0.015	0.009	0.017	0.017
	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)
Ln(capital / labor)	0.007	0.007	-0.009	0.011	0.008	0.006
	(0.007)	(0.008)	(0.016)	(0.006)	(0.007)	(0.007)
Ln(average establishment size)	0.005	0.002	0.002	0.005	0.005	0.005
	(0.002)	(0.003)	(0.004)	(0.002)	(0.002)	(0.002)
Year=1900	-0.006	-0.004	-0.002	-0.004	-0.006	-0.005
	(0.004)	(0.005)	(0.007)	(0.003)	(0.004)	(0.004)
Year=1910	-0.011	-0.008	-0.002	-0.010	-0.012	-0.009
	(0.006)	(0.007)	(0.012)	(0.006)	(0.007)	(0.006)
Observations	66,473	66,473	39,950	50,309	60,452	66,473

NOTES: The sample includes workers in the manufacturing industries, as described in the Data Appendix, for the years 1880, 1900, and 1910. The complexity measure is computed as one minus the herfindahl index of occupations, excluding managers, clerks and accountants and auditors. All regressions include industry fixed effects and city size dummies. The column specifications are: (1) Benchmark specification (sum of person weights is one for each year); (2) Standard person weights; (3) Only industries above median in terms of 1880 estaqblishment size; (4) Excludes "miscellaneous" industries; (5) Includes only industries with more than 100 observations in 1880; (6) Complexity calculated using only production workers (Craftsmen, Operatives, Laborers n.e.c.). Robust standard errors in parentheses adjusted for clustering by industry*(Year=1910).

Table 4. Employment Shares of Clerks, Managers and Computer Personnel: 1960-2000

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable:	Computer personnel				Clerks	Computer personnel	Managers
Year(s)	1970	1980	1990	2000	1960-1990	1960-1990	1960-1990
Complexity	0.0021	0.0070	0.0099	0.0174			
	(0.0007)	(0.0023)	(0.0057)	(0.0057)			
(Complexity in 1960) x (year=1980 or 1990)					-0.0650	0.0028	0.0236
					(0.0078)	(0.0011)	(0.0060)
Ln(capital/labor)					0.0002	0.0027	0.0285
					(0.0086)	(0.0014)	(0.0084)
Year=1970					-0.0095	0.0008	-0.0065
					(0.0036)	(0.0004)	(0.0030)
Year=1980					0.0245	0.0012	-0.0120
					(0.0052)	(0.0009)	(0.0053)
Year=1990					0.0108	0.0020	0.0063
					(0.0071)	(0.0012)	(0.0066)
City size dummies	No	Yes	Yes	No	No	No	No
Industry dummies	No	No	No	No	Yes	Yes	Yes
Observations	242,389	254,621	237,407	233,649	948,662	948,662	948,662

NOTES: The sample includes workers in the manufacturing industries, as described in the Data Appendix. The complexity measure is computed as one minus the herfindahl index of occupations, excluding managers, clerks and accountants and auditors. All regressions are weighted (using person-weights) and they all include state dummies. Dummies for the size of place in which the person lived were included in the years for which this information was available (1980 and 1990). Regressions (5)-(7) use normalized weights, such that the sum of weights in each year is equal to one. Robust standard errors for columns the columns (1)-(4) are clustered at the industry level and for columns (5)-(7) they are clustered at the industry*(year=1980 or 1990) level.

Table 5. Comparing USA and Mexico: 1990

	A. Summary Statistics	
	USA	Mexico
Clerks	0.112 (0.029)	0.088 (0.040)
Managers	0.104 (0.037)	0.035 (0.015)
Complexity	0.938 (0.055)	0.881 (0.113)
Industries	44	44

	B. Regression Results	
	USA	Mexico
Clerks		
Complexity	0.148 (0.021)	0.190 (0.032)
Managers		
Complexity	0.231 (0.030)	0.060 (0.008)
Observations	1,201,193	377,855

NOTES: The sample includes the workers in the manufacturing industries, according to the IPUMS International 1990 classification of industries, except miscellaneous manufacturing industries and not specified manufacturing industries. Specific occupations that were likely the result of classification error are excluded from the sample, and the industry codes have been harmonized for the US and Mexico (see Data Appendix). The complexity measure is computed as one minus the herfindahl index of occupations, excluding managers, clerks and accountants and auditors. In panel A standard deviations are in parentheses. In panels B, robust standard errors, clustered by industry, are in parentheses.

Table 6. Complexity and Different US Clerical Occupations: 1990

	A. Summary Statistics			
	All Clerks	Coordination Clerks	General Clerks	Other Clerks
	0.117	0.017	0.088	0.012
	(0.321)	(0.128)	(0.283)	(0.108)

	B. Regression Results			
	All Clerks	Coordination Clerks	General Clerks	Other Clerks
LHS Variable: Complexity	0.111	0.027	0.078	0.006
	(0.045)	(0.008)	(0.043)	(0.008)
Complexity coefficient x 1.s.d. complexity	0.0078	0.0019	0.0054	0.0004
As percent of mean:	6.7	11.3	6.2	3.8
Observations	1,281,782	1,281,782	1,281,782	1,281,782

NOTES: The sample includes the workers in the manufacturing industries, according to the IPUMS International 1990 classification of industries, which includes 82 industries. Specific occupations that were likely the result of classification error are excluded from the sample (see Data Appendix). The complexity measure is computed as one minus the herfindahl index of occupations, excluding managers, clerks and accountants and auditors. Mean industry-level complexity is 0.924 and the standard deviation is 0.070. In panel A standard deviations are in parentheses. In panels B, robust standard errors, clustered by industry, are in parentheses.

Appendix Table A1. Industries with High and Low Levels of Complexity: 1880-2000

	1880	1910	1940	1970	2000
Highest levels of complexity					
1	Blast furnaces, steel works and rolling mills	Printing, publishing, and allied industries	Ship and boat building and repairing	Ship and boat building and repairing	Ship and boat building and repairing
2	Drugs and medicines	Petroleum refining	Petroleum refining	Petroleum refining	Aircraft and parts
3	Primary nonferrous industries	Fabricated steel products	Railroad and misc transportation equipment	Blast furnaces, steel works and rolling mills	Petroleum refining
4	Printing, publishing, and allied industries	Railroad and misc transportation equipment	Aircraft and parts	Aircraft and parts	Drugs and medicines
5	Professional equipment	Agricultural machinery and tractors	Agricultural machinery and tractors	Cement, concrete, gypsum and plaster products	Railroad and misc transportation equipment
Lowest levels of complexity					
1	Cement, concrete, gypsum and plaster products	Tobacco manufactures	Footwear, except rubber	Apparel and accessories	Apparel and accessories
2	Knitting mills	Leather products, except footwear	Knitting mills	Knitting mills	Carpets, rugs, and other floor coverings
3	Petroleum refining	Footwear, except rubber	Apparel and accessories	Footwear, except rubber	Leather products, except footwear
4	Canning and preserving fruits, vegetables, and seafoods	Knitting mills	Leather products, except footwear	Leather products, except footwear	Paperboard containers and boxes
5	Confectionary and related products	Paperboard containers and boxes	Tobacco manufactures	Canning and preserving fruits, vegetables, and seafoods	Rubber products

NOTES: The sample includes the manufacturing industries, accoring to the IPUMS 1950 classification of industries. Industries with vague classification and industries that had fewer than 10 observations in a given year were excluded from this table.

Appendix Table A2. Education of Clerks and Other Workers in Manufacturing

	1940				1980	
	Age 50 and over		No age restriction		No age restriction	
	Clerks	Non-clerks	Clerks	Non-clerks	Clerks	Non-clerks
Education level:						
No high school	0.46	0.77	0.20	0.57	0.04	0.14
Some high school	0.18	0.10	0.20	0.21	0.12	0.22
High school graduates	0.24	0.08	0.44	0.16	0.49	0.37
Some college	0.07	0.03	0.11	0.04	0.27	0.17
College graduates	0.05	0.03	0.05	0.03	0.07	0.11

NOTES: The sample includes workers in manufacturing industries, and it is described in detail in the Data Appendix. The means are calculated using census person weights.

Appendix Table A3. Wages of Clerks in Manufacturing

	Wage premium of clerks relative to production workers		Annual earnings of clerks relative to production workers	
	Men	Women	Men	Women
		A. 1890-1939		
1890				1.85
1895			1.69	1.94
1909			1.65	1.96
1914			1.70	2.07
1919			1.20	1.53
1924			1.10	1.40
1929			1.13	1.53
1939			1.15	1.56
		B. 1939-1999		
1939	0.89	1.12	1.12	1.70
1949	0.97	0.99	1.02	1.30
1959	0.90	0.91	1.10	1.23
1969	0.89	0.86	1.10	1.21
1979	0.87	0.87	1.06	1.19
1989	0.82	0.80	0.97	1.01
1999	0.77	0.74	0.91	0.89

NOTES: The estimates in Panel A are taken from Table 5 in Goldin and Katz (1995), and they are based in part on sources that are not nationally representative. The estimates in Panel B are based on the author's calculations from IPUMS data. The earnings premium for clerks is defined as the ratio of the median annual earnings of clerks and non-clerks in manufacturing industries.

Appendix Table A4. USA and Mexican Industry Codes, 1990

Industry name	US industry codes	Mexico industry codes
Meat and Fish products	100	31001 31002
Dairy products	101	31003 31004
Fruit and vegetable products	102	31005
Grain mill products	110	31008 31006
Bakery products	111	31007 31011
Sugar and confectionery products	112	31015 31016 31017
Beverage industries	120	31031 31032 31033 31012
Misc. food preparations and kindred products	121	31013 31014 31021 31022
Food industries, n.s.	122	31099
Tobacco manufactures	130	31041
Textile mill products	140 141 142 150	31101 31102
Apparel and accessories	151 132	31105 31106
Fabricated textile products	152	31104 31111 31103 31199
Paper products	160 161 162	32001 32099
Publishing and printing	171 172	32011
Plastics, synthetics, and resins	180	32105
Drugs	181	32111
Soaps and cosmetics	182	32113 32114
Paints, varnishes, and related products	190	32112
Agricultural chemicals	191	32104
Industrial and miscellaneous chemicals	192	32103 32115 32116
Petroleum refining	200	32121
Miscellaneous petroleum and coal products	201	32131 32132 32101 32102
Rubber products	210 211	32141
Plastics products	212	32151 32152
Leather products	220 222	31201
Footwear, except rubber and plastic	221	31211
Wood products, other than furniture	230 231 232 241	31301 31321 31399
Furniture and fixtures	242	31311
Glass and glass products	250	32211
Construction materials	251 252	32221 32202 32222
Pottery and related products	261	32201
Misc. nonmetallic mineral and stone products	262	32299
Primary Iron and steel industries	270 271	32301 32401
Primary non-ferrous metal industries	272	32311
Other primary metal industries	280	32399
Fabricated metal products	281 282 290 291 292 300 301	32402 32403 32404
Misc. mechanical machinery	310 311 312 320 331 332	32441
Office and accounting machines	321 322	32413
Professional instruments and equipment	371 372 380 381	32411 32412
Household appliances	340	32423
Misc. electronic equipment	341 342 350	32422
Motor vehicles and motor vehicle equipment	351	32431 32421
Misc. transportation equipment	352 360 361 362 370	32432

NOTES: Data is from IPUMS International, 1990 Classification of industries for USA and Mexico

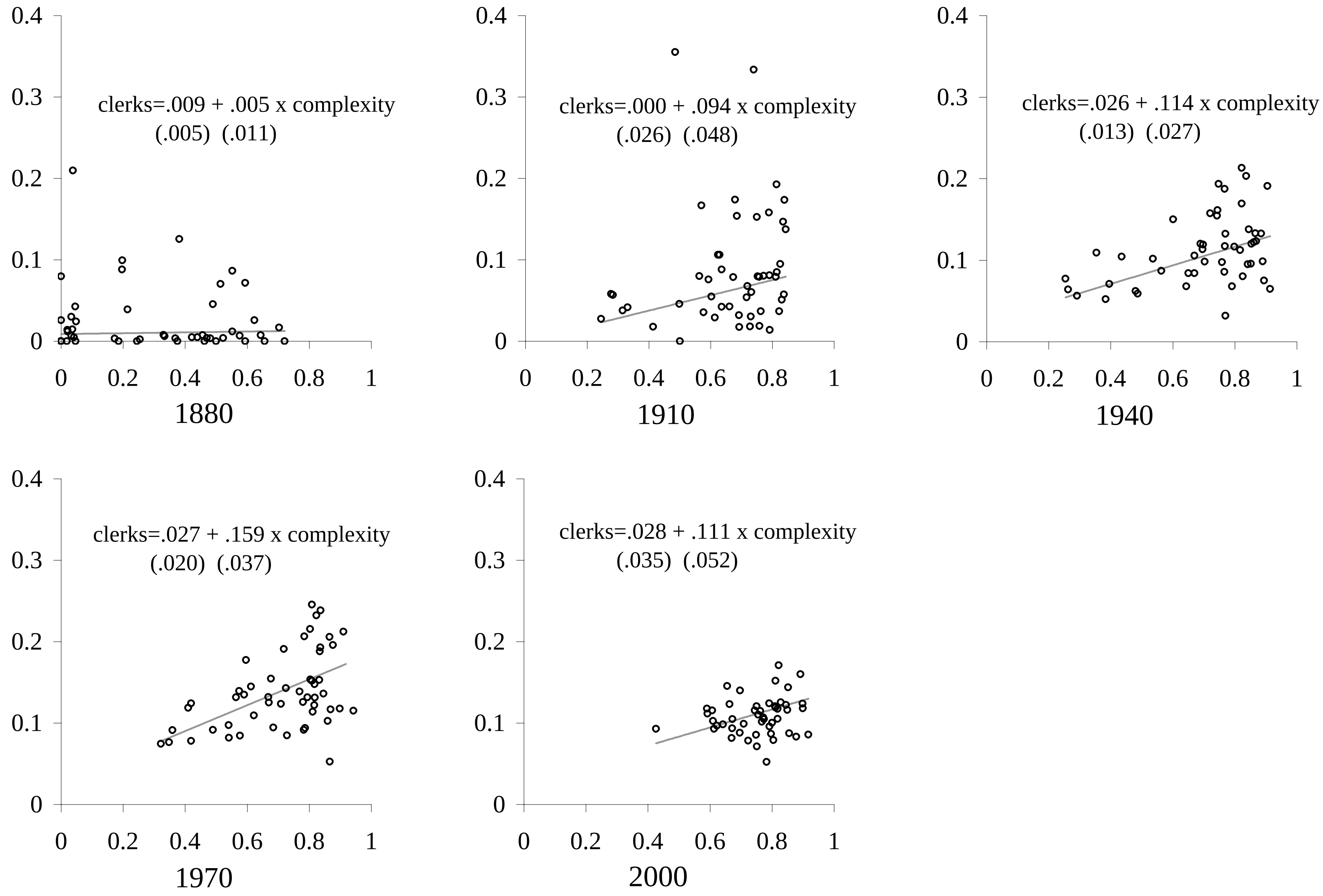

Figure 1. Percentage of clerks in workforce as a function of complexity: 1880-2000. Source: author's calculations.

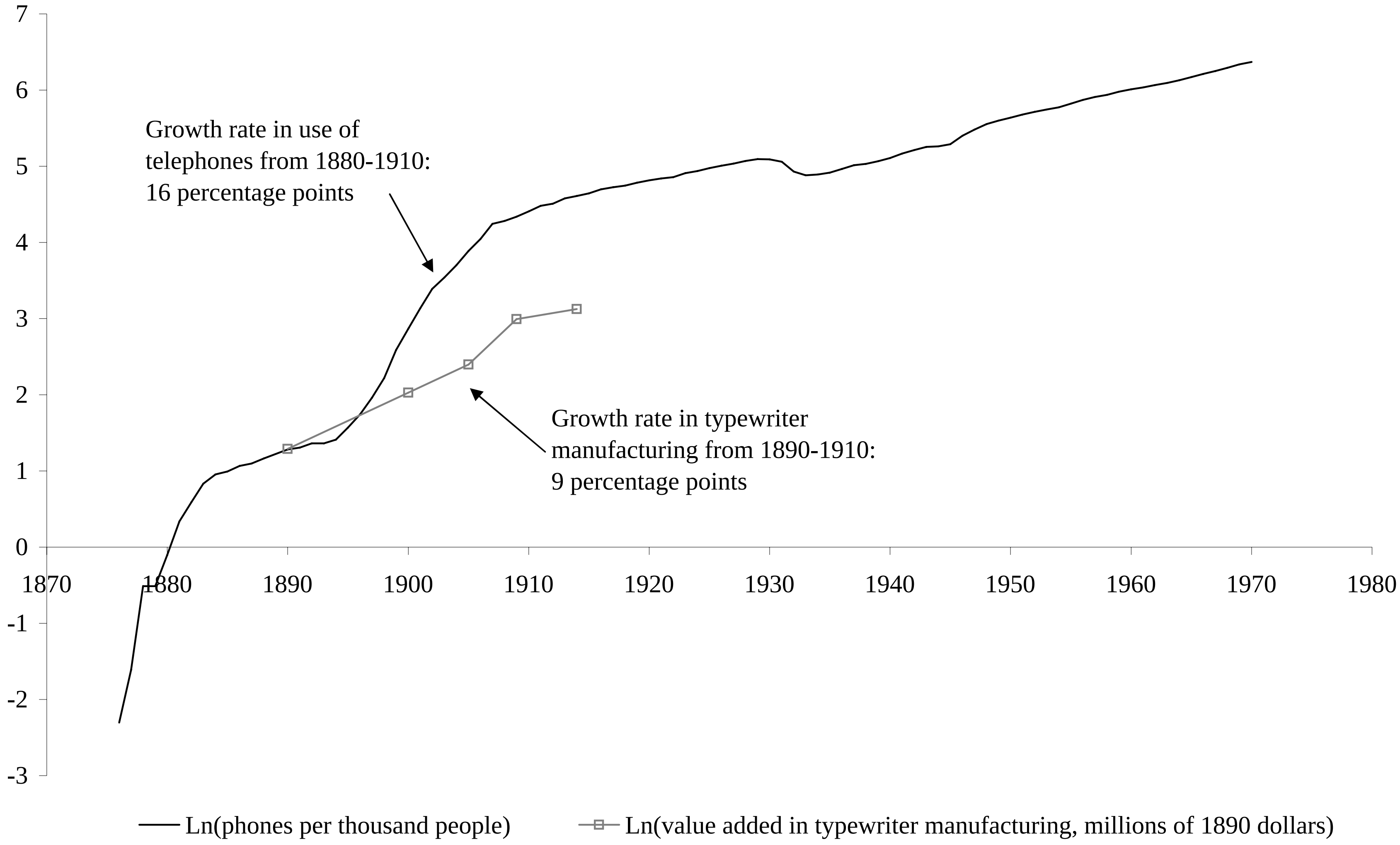

Figure 2. Diffusion of information technology equipment during the early IT revolution. Source: author's calculations based on the Census of Manufactures.

Figure 3. Fraction of white-collar workers in manufacturing, by occupational class. Source: author's calculations.

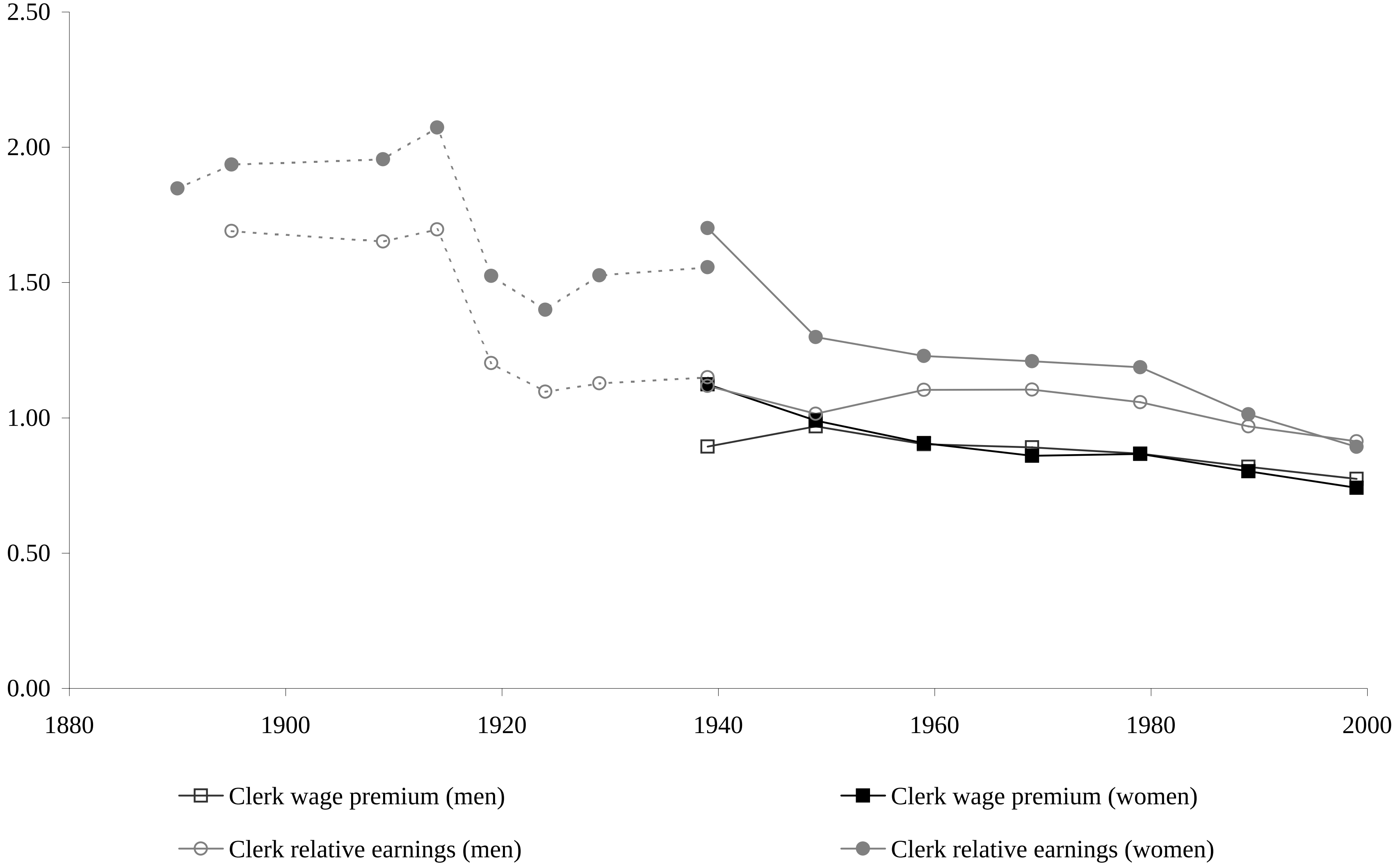

Figure 4. Wages of clerks relative to production workers in manufacturing: 1880-2000. Broken lines denote data from before 1940, which are not nationally representative. Sources: see Appendix Table A2.