Evidence on Wage Inequality, Worker Education, and Technology

Christopher H. Wheeler

The rise in U.S. wage inequality over the past two decades is commonly associated with an increase in the use of "skill-biased" technologies (e.g., computer equipment) in the workplace, yet relatively few studies have attempted to measure the direct link between the two. This paper explores the relationship among inequality, worker education levels, and workplace computer usage using a sample of 230 U.S. industries between 1983 and 2002. The results generate two primary conclusions: First, this rising inequality in the United States has been caused predominantly by increasing wage dispersion within industries rather than between industries. Second, within-industry inequality is strongly tied to both the frequency of computer usage among workers and the fraction of total employment with a college degree. Both results lend support to the idea that skill-biased techno-logical change has been an important element in the rise of U.S. wage inequality.

The rapid rise of U.S. wage inequality in recent decades has produced a sizable literature both documenting the empir-ical trends and theorizing about their causes.[1] The main empirical findings can be sum-marized by three basic patterns. First, the overall distribution of hourly and weekly earnings across all workers in the economy has grown wider. Second, consistent with this rise, the wage gaps between workers with different levels of educa-tion, especially between college graduates and workers with no more than a high school diplo-ma, have also increased. This rise in "between-education-group" earnings disparity, however, accounts for only a modest fraction of the rise in overall wage dispersion because of the third pattern: The variance of wages among workers with the *same* level of education has also grown.[2]

To explain these patterns, a variety of theories have been advanced, including those stressing the growth of international trade, changes in institu-tions (e.g., declining unionization and real mini-mum wage), rising immigration, and technological change. Growing levels of imports into the United States, for instance, may have hit workers in trade-sensitive industries (e.g., textiles and apparel) particularly hard as domestic labor demand and, consequently, wages have dropped.[3] Rising immi-gration since the 1960s may also have contributed to these trends by increasing the supply of less-skilled workers in the U.S. labor market (Borjas, Freeman, and Katz, 1997). In addition, because unionization is often associated with wage com-pression (Fortin and Lemieux, 1997), declining rates of union membership in the United States may have contributed to rising earnings disparity.[4]

---

[1] See Levy and Murnane (1992) and Acemoglu (2002) for surveys.

[2] These basic patterns have also been observed for a number of Organisation for Economic Co-operation and Development (OECD) countries, although wage inequality in the United States remains higher than that of most other developed economies. See Blau and Kahn (1996).

[3] The Bureau of Labor Statistics produces the *Occupational Outlook Handbook*, which offers predictions about job and wage growth in various sectors, including Textiles and Apparel, given (among other things) trends in international trade. The most recent edition can be found at www.bls.gov/oco/home.htm.

[4] In addition to Fortin and Lemieux (1997), Topel (1997) and Johnson (1997) offer surveys of several prominent theories of wage inequality.

While there remains some disagreement as to the significance of each of these mechanisms, a general consensus has formed around one particular theory: skill-biased technological change (SBTC).[5] The hypothesis is quite simple. Over the past several decades, the supply of highly educated workers in total employment has grown. In 1950, for example, 17 percent of U.S. workers had some education at the college level. By 1990, 57 percent did.[6] As a result of this increase, the return to investing in technologies that complement the skills of these highly educated workers—the most commonly cited example of which is information technology—also rose because the search costs involved in finding and hiring "skilled" labor declined.[7] Accordingly, recent technological change has served to boost the wages of skilled workers while depressing the employment opportunities and earnings of the less-skilled. As noted by Acemoglu (1999), if "skills" are positively but imperfectly associated with educational attainment, this mechanism can lead to larger between-education-group gaps as well as greater inequality *within* education groups.

Although there has been a host of evidence documented on the general topic of technological change, skill distributions, and inequality, there are (at least) two issues that remain unresolved. First, to what extent is rising inequality a within- or between-sector phenomenon? That is, does the SBTC argument imply that inequality is driven by growing differentials across workers in different sectors or by growing gaps within the same sector? Caselli (1999), for example, reports that, between 1975 and 1991, the variance of (equipment) capital-labor ratios across 450 four-digit manufacturing sectors rose sharply in the United States. Because capital-labor ratios tend to correlate positively with both wages and the fraction of highly skilled workers in total employment, Caselli interprets this trend as evidence that SBTC has been highly

variable across sectors: Some industries have adopted advanced technologies and hired educated workers; others have chosen to utilize skill-*un*biased technologies and less-educated workers. This result suggests that SBTC has driven inequality higher through a between-industry channel. A qualitatively similar result is reported by Acemoglu (1999), who finds that the fraction of workers in the United States holding jobs (defined by 174 industry-occupation cells) in the bottom and top tails of the distribution of average hourly pay increased between 1983 and 1993. This finding, he concludes, indicates that workers have increasingly been sorted into "good" and "bad" jobs, which further suggests that rising inequality has been the product of growing between-sector dispersion.

On the other hand, many studies of inequality (e.g., Katz and Murphy, 1992; Juhn, Murphy, and Pierce, 1993; Card and DiNardo, 2002) suggest that, even after accounting for observable differences across workers (including their industries of employment), the dispersion in their wage earnings has risen markedly. Such evidence suggests that within-sector differences must also be an important aspect of rising dispersion. However, it remains unclear just how important these two elements have been in explaining the overall increase in earnings disparity.

Second, much of the evidence on technological change and wage dispersion tends to be indirect. That is, in spite of the popularity of the SBTC hypothesis, surprisingly little research has directly examined the association between inequality and the extent to which computer equipment (or any other "advanced" technology) is used in production.[8] Most studies have either connected average wage earnings to the use of computers and other sophisticated technologies (Krueger, 1993; Doms, Dunne, and Troske, 1997) or explored the relationship between the adoption of information technology and the distribution of worker skill (Berman, Bound, and Griliches, 1994; Doms, Dunne, and Troske, 1997; Autor, Katz, and Krueger, 1998).

---

[5] As noted by Card and DiNardo (2002), the SBTC explanation is far from complete. Nonetheless, as suggested by Johnson (1997), there is wide agreement that it has been an important determinant of rising inequality.

[6] These statistics are reported by Wheeler (2004).

[7] See Autor, Levy, and Frank (2003) for an empirical analysis of why computers are considered skill-biased.

[8] Dunne, Foster, and Troske (2004) is a notable exception. However, the focus of that paper is the U.S. manufacturing sector rather than the entire private U.S. economy, which I examine here.

This paper seeks to address both of these issues. To this end, I perform two exercises. In the first, I use annual data from the Current Population Survey (CPS) over the period 1983-2002 to evaluate the degree to which rising wage dispersion in the entire private U.S. economy can be attributed to growing dispersion within industries as opposed to between them. In the second, I look at the relationship between, on the one hand, a variety of inequality measures within individual industries and, on the other, the distribution of educational attainment and the extent of computer usage among workers employed in those industries. Computer usage, I assume, provides a direct measure of SBTC; educational attainment provides an indirect measure.[9]

To summarize briefly, the results indicate that the rising U.S. wage inequality has been driven primarily by growing dispersion among workers within the same industry rather than between industries. In each year, more than 75 percent of the variance of hourly earnings can be attributed to within-industry variation. More importantly, this fraction has grown steadily over time, suggesting that the majority of the rise in overall wage variance is due to increasing wage disparity among workers within the same industry. When I turn to the analysis of inequality within industries, I find that wage dispersion—measured in a variety of ways—is positively associated with both the fraction of college-educated workers and the extent of computer usage. These results, I conclude, offer some support for the skill-biased technological change argument.

DATA

Sources

The majority of the worker-level data used in this paper are derived from the Merged Outgoing Rotation Group (MORG) files of the CPS for each year between 1983 and 2002. These files are constructed by combining the individuals from each month's CPS who are in their final month (i.e., fourth or eighth) of interview and are, conse-

quently, asked about their labor earnings. In an effort to focus on individuals of prime working age, I limit the sample to workers between the ages of 18 and 65.

I perform two sets of calculations from the MORG files. First, I compute educational attainment distributions and union membership rates for a collection of more than 200 industries, which correspond to an approximately three-digit (Standard Industrial Classification [SIC]) level of aggregation. To maximize the number of observations used for these computations, I use all individuals for whom an industry is identified and who report positive weekly earnings. Doing so produces a sample of 2,693,370 observations across the 20 years.

The second set of calculations involves a variety of earnings inequality measures based on hourly wages. Here, I further limit the sample to white males who report working at least 30 hours per week so that the sample consists entirely of workers with a strong attachment to the labor force (i.e., their primary activity is work). Doing so eliminates the need to account for the influence of race and gender on earnings and, thus, the attendant inequality. It also reduces the effects of part-time workers whose presence in the workforce from one year to the next may be heavily influenced by the business cycle. I further confine the sample to hourly wages between $2.60 and $150 (in year-2000 dollars) to remove any remaining outlier observations. These sample selection criteria are reasonably standard in the wage inequality literature (e.g., Katz and Murphy, 1992; Juhn, Murphy, and Pierce, 1993; Card and DiNardo, 2002). In all, 1,156,715 observations are used in the inequality calculations.

The industry coverage includes the entire private sector. As noted, industries are mostly defined at a three-digit (SIC) level of aggregation, although some two-digit sectors and combinations of either three- or four-digit sectors are also represented.[10] For example, Coal Mining (CPS

---

[9] If SBTC is indeed a function of the distribution of skill, the fraction of highly educated workers should capture (at least to a significant degree) the extent of SBTC within an industry.

[10] A major (one-digit) sector is also included, Construction. To provide some sense of the differences between two-, three-, and four-digit industries, Pharmaceutical Preparations and Medicinals and Botanicals are four-digit sectors that belong to the three-digit sector Drugs, which, in turn, is included in the two-digit sector Chemicals and Allied Products.

industry code 41) and Air Transportation (code 421) are two-digit sectors; Hardware Stores (code 581), Drugs (181), and Advertising (code 721) are three-digit sectors; Glass and Glass Products (code 250) is a combination of three-digit sectors; and Primary Aluminum (code 272) is a collection of four-digit sectors. A total of 230 industries are identified over the 20-year period.

To quantify computer usage rates within each industry, I use the October supplements to the CPS for the years 1984, 1989, 1993, and 1997. In these supplements, individuals were asked about their computer use at work, including (for some of the years) the types of tasks performed using this equipment. Computer usage rates are calculated as frequencies of positive responses to the question: "Do you directly use a computer at work?" Note that these are the same data used by Krueger (1993) in his study of the effect of information technology on wages and by Autor, Katz, and Krueger (1998) in their analysis of computer usage and skill distributions. Here, too, to maximize the number of observations used to estimate computer usage within detailed sectors, the calculations incorporate all workers for whom an industry of employment can be identified. Additional details about the construction of the final data set appear in the appendix.

Some Trends

As noted in the introduction, three broad features characterize the evolution of the U.S. wage distribution in recent decades: rising overall dispersion, widening gaps between workers with different levels of education, and increasing dispersion among workers with the same levels of education. All three are evident in the CPS data examined here.

In 1983, for example, the 90th percentile of the overall hourly wage distribution was roughly 1.9 times as large as the median. By 2002, it was 2.2 times as large. Figure 1 shows a similar result using an alternative measure of overall wage dispersion, the variance of log hourly wages.[11]

Although there have been years in which the variance has decreased, the general trend has clearly been upward, rising 19 percent between 1983 and 2002.

Between-education-group gaps also exhibit an upward trend. High school graduates in the sample, for instance, earned roughly 16 percent more than high school dropouts in 1983.[12] In 2002, they earned 27 percent more. What is even more striking, however, is the gap at the top end of the educational attainment distribution. Figure 2 plots the evolution of the wage premium earned by college graduates relative to workers with only a high school diploma. This wage differential increased from an average of 54 percent in 1983 to 73 percent in 2002.[13]

To see that earnings differentials among workers with the *same* levels of education have also grown over this period, consider Figure 3, which plots the variance of the residuals following a regression of log hourly wages on education and experience.[14] Based on the calculations, the variance of these "residual earnings"—which is often interpreted as the degree of spread in the earnings distribution of workers with the same observable levels of skill—rose by nearly 20 percent over this 20-year period.

BETWEEN-INDUSTRY VERSUS WITHIN-INDUSTRY INEQUALITY
Overall Wages

To assess the degree to which rising U.S. earnings dispersion has been a between- or within-industry phenomenon, I consider the following straightforward decomposition. Given a sample

---

[11] Wages are usually expressed in logarithms in studies of labor earnings. Doing so facilitates both the computation and interpretation of the results (see Card, 1999).

[12] These figures are based on the coefficients from year-specific regressions of log hourly wages on four educational attainment dummies (no high school, some high school, high school, some college, college); a quartic polynomial in potential experience; and indicators for marital status, union membership, metropolitan status, and Census division of residence.

[13] Figure 2 plots "log point" differences (i.e., the difference between the log wages of one group and the log wages of another). To derive percentages, simply calculate $(exp(x) - 1)$, where x is the log point difference.

[14] More precisely, these residuals are derived from the same regression as that described in footnote 12.
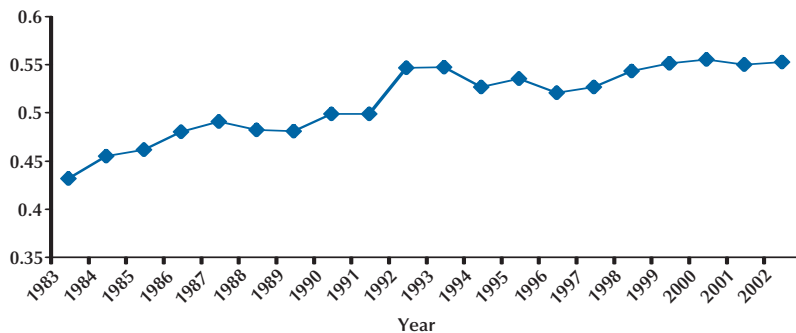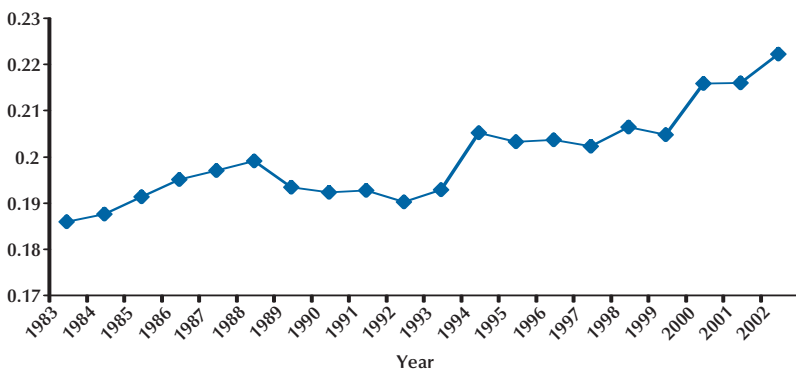
Figure 1

Overall Wage Variance of Log Hourly Wages
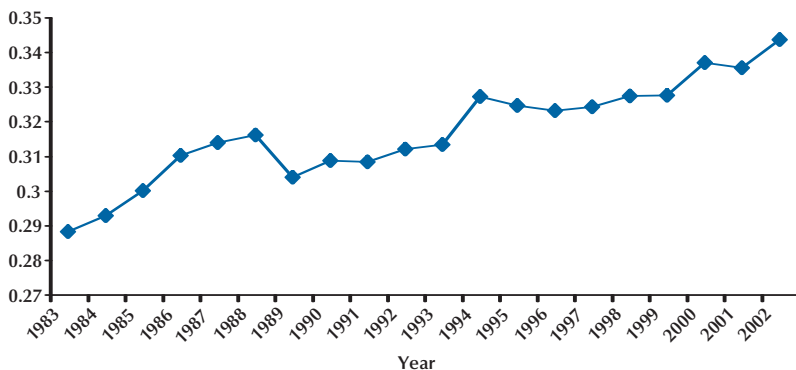
Figure 2

College Graduates Relative to High School Graduates

Figure 3

Residual Wage Variance of Log Hourly Wages

Table 1

Overall Inequality Decomposition

Year	Total wage variance	Within-industry component	Between-industry component
1983	0.288	0.222 (77.1)	0.066 (22.9)
1984	0.293	0.228 (77.9)	0.065 (22.1)
1985	0.3	0.23 (76.8)	0.069 (23.2)
1986	0.31	0.237 (76.6)	0.073 (23.4)
1987	0.313	0.24 (76.6)	0.073 (23.4)
1988	0.316	0.242 (76.8)	0.073 (23.2)
1989	0.304	0.233 (76.8)	0.07 (23.2)
1990	0.308	0.237 (76.9)	0.071 (23.1)
1991	0.308	0.235 (76.5)	0.072 (23.5)
1992	0.312	0.237 (76.1)	0.075 (23.9)
1993	0.313	0.24 (76.7)	0.073 (23.3)
1994	0.327	0.259 (79.3)	0.068 (20.7)
1995	0.324	0.259 (79.9)	0.065 (20.1)
1996	0.323	0.256 (79.2)	0.067 (20.8)
1997	0.324	0.258 (79.8)	0.066 (20.2)
1998	0.327	0.261 (79.7)	0.066 (20.3)
1999	0.327	0.262 (80)	0.065 (20)
2000	0.337	0.266 (79)	0.071 (21)
2001	0.335	0.267 (79.7)	0.068 (20.3)
2002	0.343	0.275 (80.2)	0.068 (19.8)

NOTE: Between- and within-industry components of total variance in log hourly wages as defined by equation (3). Percentages of total variance accounted for by each component are reported in parentheses. The final column is calculated by dividing the annual changes in the within-industry component by the corresponding changes in total variance.

of workers, the variance of the hourly wage distribution in year t, V_t, can be estimated as

$$(1) \qquad V_t = \frac{1}{N_t} \sum_{i=1}^{I_t} \sum_{j=1}^{N_{i,t}} (w_{j,i,t} - \bar{w}_t)^2,$$

where $w_{j,i,t}$ is the wage of worker j of industry i, $\bar{w}_t$ is the overall mean wage, $N_{i,t}$ is the number of workers in industry i, I_t is the number of industries, and N_t is the total number of workers, $\Sigma_i N_{i,t}$, all for the year t. If we rewrite this expression as

$$(2) \qquad V_t = \frac{1}{N_t} \sum_{i=1}^{I_t} \sum_{j=1}^{N_{i,t}} (w_{j,i,t} - \bar{w}_{i,t} + \bar{w}_{i,t} - \bar{w}_t)^2,$$

where $\bar{w}_{i,t}$ denotes the mean wage among workers of industry i, the variance of the wage distribution can be expressed as the sum of two terms[15]:

$$(3)$$

$$V_t = \frac{1}{N_t} \sum_{i=1}^{I_t} \sum_{j=1}^{N_{i,t}} (w_{j,i,t} - \bar{w}_{i,t})^2 + \frac{1}{N_t} \sum_{i=1}^{I_t} \sum_{j=1}^{N_{i,t}} (w_{i,t} - \bar{w}_t)^2.$$

The first, given by the sum of squared deviations of individual wages from their industry means, can be interpreted as a "within-industry" component of wage dispersion. The second, which is constructed from the sum of squared deviations of the industry means from the overall mean, can be viewed as a "between-industry" component. By calculating these two pieces, we can gain some insight into the extent to which rising wage inequality in the United States over the past two

[15] The derivation of (3) is sketched in the appendix.

Table 2

Residual Inequality Decomposition

Year	Total wage variance	Within-industry component	Between-industry component
1983	0.186	0.159 (85.7)	0.026 (14.3)
1984	0.187	0.162 (86.7)	0.025 (13.3)
1985	0.191	0.165 (86.2)	0.026 (13.8)
1986	0.195	0.167 (85.8)	0.028 (14.2)
1987	0.197	0.17 (86.4)	0.027 (13.6)
1988	0.199	0.171 (86.3)	0.027 (13.7)
1989	0.193	0.168 (87)	0.025 (13)
1990	0.192	0.168 (87.3)	0.024 (12.7)
1991	0.192	0.169 (87.7)	0.024 (12.3)
1992	0.19	0.167 (87.7)	0.023 (12.3)
1993	0.193	0.17 (88.3)	0.022 (11.7)
1994	0.205	0.184 (90)	0.021 (10)
1995	0.203	0.183 (90.2)	0.02 (9.8)
1996	0.203	0.184 (90.2)	0.02 (9.8)
1997	0.202	0.183 (90.7)	0.019 (9.3)
1998	0.206	0.187 (90.9)	0.019 (9.1)
1999	0.205	0.187 (91.5)	0.017 (8.5)
2000	0.216	0.196 (90.8)	0.02 (9.2)
2001	0.216	0.197 (91.4)	0.019 (8.6)
2002	0.222	0.204 (92)	0.018 (8)

NOTE: Between- and within-industry components of total variance in residual log hourly wages (after a regression on education and experience) as defined by equation (3). Percentages of total variance accounted for by each component are reported in parentheses. The final column is calculated by dividing the annual changes in the within-industry component by the corresponding changes in total variance.

decades has been a between- or within-industry phenomenon.

The resulting components using overall (i.e., unconditional) log hourly wages are listed in Table 1. Most obviously, they show that the vast majority of the wage dispersion observed each year is due to earnings variation within industries. This particular result can be seen from the fractions of total variation accounted for by each component, which are reported in parentheses. In all years, the fraction of total wage variance accounted for by within-industry variation is between 75 and 80 percent. More importantly, there seems to have been a gradual rise in this fraction over time. At the beginning of the sample time frame, the within-industry component averaged approximately 77

percent of the total. By 2002, it was closer to 80 percent.

In all, the variance of log hourly wages rose from 0.29 to 0.34 between 1983 and 2002. While both the within- and between-industry components also rose over this time frame, the within-industry part accounted for nearly all—approximately 96.4 percent—of the increase in total variance.

Residual Wages

As noted previously, one of the basic features of the rise in U.S. wage inequality is the growing degree of dispersion in the wage earnings of workers with similar levels of skill (i.e., education and experience). That is, the degree of variation among,

Figure 4

Computer Usage and College Completion Rates
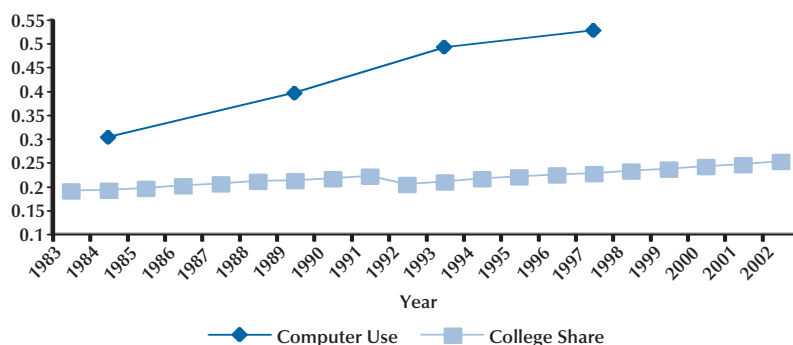
say, college graduates with 10 years of work experience has grown larger in recent decades. Has this rise in residual inequality also been primarily a within-industry phenomenon?

To consider this question, I begin by regressing log hourly wages on a vector of personal covariates, including years of education; four educational attainment indicators (no high school, some high school, high school, some college, college or more); interactions of years of education with these indicators; a quartic polynomial in potential work experience; and dummies for marital status, union membership, residence in a metropolitan area, and Census division.[16] I then collect the residuals and use them in place of actual wages, $w_{j,i,t}$, when calculating the within- and between-industry pieces in (3).

Table 2 shows the results. Qualitatively, they show precisely the same result as with overall wages. In each year, the within-industry component is by far the larger piece of overall variation, averaging between, roughly, 85 and 90 percent of the total. Additionally, there has been a gradual rise in this fraction over time—from 86 percent in 1983 to approximately 92 percent in 2002—

suggesting that the within-industry component has become more important over time.

On the whole, between 1983 and 2002, the change in within-industry residual wage variation actually exceeded the increase in total residual wage variation. To be specific, increases in within-industry residual wage variation accounted for 123.9 percent of the change in total residual variation. Hence, there was actually a net decrease in the extent of inequality across workers possessing similar characteristics but employed in different industries over these years. Evidently, whatever has driven the rise of inequality across workers— either with similar levels of education or not— has done so primarily within individual industries.

INEQUALITY, EDUCATION, AND COMPUTER USE

Baseline Results

Given that the majority of rising wage dispersion has been the result of growing differentials among workers within the same industry, I now turn to the analysis of within-industry inequality trends. Specifically, this section examines the role of skill distributions (i.e., fractions of highly educated workers in total employment) and information technology in explaining industry-specific earnings inequality.

As shown in Figure 4, the fraction of total

---

[16] These regressions are performed separately for each year. Again, the logarithmic transformation of wage earnings is standard in the labor literature as is the specification of the wage-experience profile by means of a fourth-order polynomial (e.g., Autor, Katz, and Krueger, 1998). In terms of years of schooling completed, the educational attainment categories correspond to 0-8 (no high school), 9-11 (some high school), 12 (high school), 13-15 (some college), 16+ (college).

Table 3

Education and Computer Usage Changes by Major Sector

Industry	College employment fraction change 1983-2002	Computer usage change 1984-97
Agriculture, forestry, fisheries	0.02	0.17
Mining	–0.06	0.07
Construction	–0.001	0.13
Nondurable manufacturing	0.09	0.23
Durable manufacturing	0.05	0.15
Transportation	0.055	0.19
Wholesale trade	0.04	0.18
Retail trade	0.013	0.21
Finance, insurance, real estate	0.11	0.24
Business and repair services	0.11	0.29
Personal services	0.04	0.19
Entertainment and recreation services	0.05	0.26
Professional and related services	0.05	0.32

NOTE: Changes in proportions of employees with a bachelor's degree and using a computer at work.

employment accounted for by college graduates and the fraction of workers who use a computer at work have both increased in the past two decades. In 1983, the fraction of workers with at least a bachelor's degree stood at 19 percent. By 2002, it had grown to 25 percent. Similarly, between 1984 and 1997, the fraction of workers reporting use of a computer at work increased from 30 percent to approximately 53 percent.

Table 3 shows that both of these qualitative patterns were reasonably widespread, at least in the sense that they occurred in nearly every major industrial sector. In fact, of the 13 sectors listed in the table, only 2—Mining and Construction—saw their college employment fractions decrease over the sample period. All, as it happens, witnessed increasing computer usage. Among the 228 (of 230) more detailed industry groups identified in both the beginning and ending years of the sample, the results are similar. A total of 175 increased their fraction of college graduates in total employment between 1983 and 2002, while 225 saw increases in their computer usage rates between 1984 and 1997.[17]

To what extent do these trends account for industry-specific levels of wage inequality? I attempt to draw inferences about the answer using the following statistical characterization of inequality in industry i in year t, $Ineq_{i,t}$:

$$(4) \qquad Ineq_{i,t} = \alpha + \delta_t + \beta X_{i,t} + \varepsilon_{i,t},$$

where α is an overall constant; δ_t represents a time dummy added to capture the temporal variation in inequality evident from Figures 1, 2, and 3[18]; $X_{i,t}$ is a vector of time-varying industry characteristics; and $\varepsilon_{i,t}$ is a residual. Three quantities are included in $X_{i,t}$: the fraction of workers with a bachelor's degree,[19] the fraction using a computer

---

[17] The mean (standard deviation) change in the college employment fraction is 0.04 (0.08); the change in the computer usage rate is 0.25 (0.16).

[18] I reestimated specification I of equation (4) further adding industry-specific time trends to capture differences in the temporal behavior of inequality across industries. The resulting estimates did not differ substantially from those reported here. All of the college-share coefficients were significantly positive; all but one of the union rate coefficients were significantly negative.

[19] I also considered the share of total work hours accounted for by college-educated workers instead of the college employment fraction. Since the correlation between these two variables exceeds 0.99, the results did not differ substantially from what is reported here.

at work, and the fraction who are members of a union. As suggested previously, computer usage is intended as a direct measure of SBTC, whereas the college share is used as an indirect measure. Both interpretations seem justifiable in light of the research surveyed here connecting technological adoption and the distribution of education/skill. The unionization rate is included to capture the influence of an institutional characteristic that has likely contributed to changes in industry-level inequality. Since these independent variables are calculated from the CPS micro samples, I restrict estimation to those industry-years for which at least 10 observations were available, in an effort to reduce the sampling noise inherent in each.

Because the computer usage data are available for only 4 of the 20 years, I consider three different specifications of this equation. In the first, I limit $X_{i,t}$ to the college and unionization fractions so that I am able to use all 20 years of data. Direct evidence correlating inequality with technology is then given in the second specification, which drops the college fraction but adds the computer usage rate. The third specification considers both direct and indirect measures of SBTC simultaneously by adding the college employment fraction to this second specification.

To keep the analysis as broad as possible, I examine three different categories of inequality measures: (i) overall, (ii) between-education-group, and (iii) residual. The overall measures include the variance of log hourly wages and the differences between the 90th, 50th, and 10th percentiles of the log hourly wage distribution. The between-education-group gaps are given by differences between the average log wages of college graduates and those in each of the following four categories: some college, high school, some high school, no high school. Residual inequality is given by the same statistics considered for overall inequality, where the calculations are done using the residuals following the regressions described in the subsection "Residual Wages."

In all cases, estimation proceeds by generalized least squares in which the industry-year observations are weighted by the number of CPS observations used to calculate the inequality variables. An inequality figure based on 10 observa-

tions, after all, ought to involve greater sampling error than one based on 1,000. This weighting procedure helps to account for the differential degree of noise across observations.

Results appear in Table 4. On the whole, they show that inequality tends to be significantly associated with each of the three regressors—the college employment fraction, the computer usage rate, and the extent of union membership. The unionization rate, as expected, enters negatively in 34 (of 36) instances (significantly in 30), suggesting that decreasing unionization is an important piece of the rise in nearly every wage gap considered. Such a result, of course, reinforces the general view established in the inequality literature that the decline in union activity in the United States has been a major element in the rise of earnings disparity.[20]

The two measures of SBTC, by contrast, both enter positively in nearly every case. Indeed, none of the college fraction coefficients and only one of the estimated computer usage coefficients are negative (albeit statistically insignificant). What is more, of the 24 coefficients for each of these two quantities, a large number are statistically important: 20 of the college fraction coefficients and 16 of the computer usage coefficients. The majority of the insignificant coefficients, incidentally, emerge from specification III in which both of these regressors appear. Quite possibly, the lack of significance in these cases derives from the strong correlation between these two variables.[21]

Are the estimated associations *economically* important? Just focusing on the overall 90-10 wage difference, the point estimates suggest that a 1-standard-deviation increase in either the college employment fraction or the computer usage rate is accompanied by a 4- to 7-percentage-point increase in this differential.[22] When evaluated at

---

[20] Blau and Kahn (1996), for instance, find that differences in labor market institutions (e.g., unionization) account for a large part of the difference between inequality in the United States and that of nine other OECD countries.

[21] The correlation between the college employment fraction and the frequency of computer usage is 0.61, which is consistent with the findings of Autor, Katz, and Krueger (1998).

[22] Summary statistics for the variables used in the inequality regressions are reported in Table A1 of the appendix.

Table 4

Baseline Inequality Regressions

Dependent variable	Specification	College fraction	Computer rate	Union rate	R^2
Overall variance	I	0.18* (0.01)	—	–0.17* (0.008)	0.38
	II	—	0.11* (0.01)	–0.23* (0.02)	0.27
	III	0.1* (0.03)	0.06* (0.02)	–0.22* (0.02)	0.31
Overall 90-10 difference	I	0.53* (0.03)	—	–0.52* (0.03)	0.35
	II	—	0.31* (0.05)	–0.66* (0.07)	0.24
	III	0.31* (0.1)	0.17* (0.07)	–0.63* (0.07)	0.29
Overall 90-50 difference	I	0.1* (0.02)	—	–0.55* (0.02)	0.33
	II	—	0.002 (0.03)	–0.62* (0.04)	0.25
	III	0.02 (0.06)	–0.006 (0.04)	–0.64* (0.05)	0.26
Overall 50-10 difference	I	0.43* (0.02)	—	0.03 (0.02)	0.26
	II	—	0.31* (0.04)	–0.04 (0.06)	0.2
	III	0.29* (0.05)	0.18* (0.04)	0.005 (0.06)	0.28
College–no high school	I	0.58* (0.05)	—	–0.002 (0.04)	0.19
	II	—	0.42* (0.06)	–0.08 (0.07)	0.19
	III	0.15 (0.16)	0.35* (0.1)	–0.06 (0.08)	0.2
College–some high school	I	0.5* (0.04)	—	–0.14* (0.02)	0.21
	II	—	0.3* (0.04)	–0.16* (0.04)	0.2
	III	0.23* (0.13)	0.2* (0.08)	–0.13* (0.04)	0.2
College–high school	I	0.17* (0.02)	—	–0.18* (0.02)	0.09
	II	—	0.12* (0.04)	–0.2* (0.05)	0.09
	III	0.13 (0.09)	0.06 (0.06)	–0.18* (0.05)	0.09
College–some college	I	0.09* (0.02)	—	–0.11* (0.02)	0.08
	II	—	0.054* (0.027)	–0.14* (0.03)	0.07
	III	0.07 (0.07)	0.02 (0.04)	–0.13* (0.03)	0.07
Residual variance	I	0.11* (0.008)	—	–0.12* (0.006)	0.34
	II	—	0.05* (0.01)	–0.17* (0.02)	0.21
	III	0.08* (0.03)	0.01 (0.02)	–0.16* (0.02)	0.26
Residual 90-10 difference	I	0.32* (0.02)	—	–0.41* (0.02)	0.32
	II	—	0.16* (0.04)	–0.52* (0.06)	0.18
	III	0.23* (0.07)	0.05 (0.05)	–0.5* (0.06)	0.23
Residual 90-50 difference	I	0.16* (0.02)	—	–0.26* (0.01)	0.26
	II	—	0.08* (0.02)	–0.32* (0.04)	0.15
	III	0.12* (0.05)	0.03 (0.03)	–0.32* (0.04)	0.19
Residual 50-10 difference	I	0.15* (0.01)	—	–0.15* (0.01)	0.21
	II	—	0.07* (0.02)	–0.2* (0.03)	0.11
	III	0.11* (0.04)	0.03 (0.03)	–0.19* (0.03)	0.14

NOTE: Coefficients from estimation of (4). Specification I uses annual data 1983-2002. Specifications II and III use only data from 1984, 1989, 1993, and 1997. Each regression includes year dummies. Heteroskedasticity-consistent standard errors are reported in parentheses; * denotes significance at the 10 percent level.

the mean value in dollar terms, this figure represents an increase of approximately $0.75 to $1.31 in the 90-10 differential. Such a magnitude is roughly one-quarter of the standard deviation of the 90-10 differentials in the sample.

Looking at the between-education-group gaps, there are some sizable associations here too. A 1-standard-deviation increase in the college fraction, for example, correlates with a 3- to 7-percentage-point ($0.27 to $0.62) increase in the college–some high school gap and a 2- to 2.5-percentage-point ($0.14 to $0.18) increase in the college–high school gap. For computer usage, a 1-standard-deviation increase is associated with a 4.5- to 7-percentage-point ($0.40 to $0.62) rise in the college–some high school difference and a 1.5 to 2.5 percentage point ($0.11 to $0.18) increase in the college–high school gap. Given sample standard deviations of, respectively, 36 and 30 percentage points for the college–some high school and college–high school gaps, these correlations are far from trivial.

There is, to be sure, some variation in the coefficient magnitudes and statistical significance of these two variables across the inequality measures. In particular, they tend to have the largest associations with those measures involving the position of the bottom of the distribution relative to the middle and top, at least when considering the overall and between-education-group gaps. For example, for the overall percentile differences, the coefficients are much larger for the 50-10 differentials than they are for the 90-50 differentials. The positive associations between the college and computer usage fractions and the overall 90-10 differential, therefore, clearly seem to be driven by the bottom half of the wage distribution.

A similar result can be inferred from the between-education-group measures, which show larger coefficients on the two SBTC variables when considering the two "top-bottom" inequality variables (college–no high school, college–some high school) than when looking at the two "top-middle" variables (college–high school, college–some college). This pattern is consistent with the idea that workers at the bottom ends of the wage and educational attainment distributions were the hardest hit by new technologies (or, at least, benefited the least from them).

Endogeneity Considerations

Although the regressors have been treated as exogenous thus far, it is possible that one in particular may be endogenous with respect to inequality: education. The fraction of an industry's workers with a college degree may, for example, be an increasing function of the relative returns paid to these workers. Hence, a rise in the college–high school gap could increase the college fraction—either by attracting more college graduates or driving away high school graduates, depending on what causes the gap to increase—which would bias the estimated coefficient on the college fraction upward.

In an effort to address this possibility, I consider the following simple exercise. I regress the annual changes in an industry's college employment fraction on the initial levels of inequality and a set of year dummies.[23] I then make inferences about the extent to which inequality influences the college fraction by examining the coefficients on inequality. A significant coefficient, naturally, would suggest that inequality levels have a non-negligible influence on the educational mix of workers.

The first column of figures in Table 5 shows the results, which have a nearly uniform lack of significance of initial inequality in explaining subsequent changes in industry-specific college fractions: Only 2 of the 12 coefficients are significant. These results seem to cast some doubt on the notion that an industry's college employment fraction is endogenous with respect to inequality.

Of course, because this specification may not adequately capture the response of education to changes in inequality, I also consider an alternative in which changes in the college employment fraction are regressed on one lag of the change in inequality (i.e., the change of an industry's college fraction between 2000 and 2001 is regressed on the change in its level of inequality between 1999 and 2000). Hence, instead of correlating subsequent changes in education with initial *levels* of inequality, this equation estimates how changes in education are associated with recent *changes*

---

[23] As before, I restrict these regressions to industry-year observations based on at least 10 observations for the inequality calculations.

Table 5

Education as a Function of Inequality

	Specification	
Measure	Initial levels	Lagged differences
Overall variance	–0.0007 (0.009)	–0.02* (0.01)
Overall 90-10 difference	0.0002 (0.0003)	–0.008* (0.005)
Overall 90-50 difference	–0.0003 (0.005)	–0.007* (0.005)
Overall 50-10 difference	0.0008 (0.005)	–0.005 (0.007)
College–no high school	–0.002 (0.002)	–0.006* (0.002)
College–some high school	0.007* (0.003)	0.001 (0.002)
College–high school	0.005 (0.004)	0.001 (0.002)
College–some college	0.005 (0.005)	–0.002 (0.004)
Residual variance	0.005 (0.01)	0.005 (0.02)
Residual 90-10 difference	0.003 (0.005)	0.003 (0.005)
Residual 90-50 difference	0.01* (0.006)	0.003 (0.006)
Residual 50-10 difference	–0.004 (0.01)	0.003 (0.008)

NOTE: Coefficients on inequality, in both initial levels and lagged first differences, from regressions of the annual change in an industry's college fraction on inequality. Regressions also include year dummies. Heteroskedasticity-consistent standard errors are reported in parentheses; * denotes significance at the 10 percent level.

in inequality. Those results appear in the second column of figures in Table 5.

Here, interestingly, a greater number of the coefficients—4 of the 12—are significantly non-zero at conventional levels (i.e., at least 10 percent). However, of these, all are negative, indicating that increases in an industry's inequality tend to be followed by decreases in its college employment fraction. This particular result implies that, if anything, the coefficients listed in Table 4 may actually be biased downward (i.e., toward zero) and, thus, understate the association between education and inequality. Although certainly not definitive, I take this evidence as *suggesting* that endogenous education does not pose a significant problem for the qualitative interpretation of the results.

Alternative Specifications

This section considers two alterations of the analysis described here. In the first, I look at the possibility that the dispersion of computer usage, rather than the mean usage rate, influences the degree of wage inequality. Indeed, there could very

well be a nonlinear relationship between the computer usage rate and the degree of spread in the wage distribution. It may be, for instance, that the relationship is positive at low levels of computer usage, but negative as the usage rate closes in on unity.

The second augments the regression considered in (4) with industry-specific fixed effects:

$$(5) \qquad Ineq_{i,t} = \alpha + \delta_t + \mu_i + \beta X_{i,t} + \varepsilon_{i,t},$$

where μ_i is a constant element influencing the degree of earnings inequality in industry i. Doing so controls for all time-invariant industry characteristics, unobserved or otherwise, that may influence inequality and, thus, eliminates any bias resulting from the omission of these characteristics.[24]

---

[24] One could also treat the industry-specific terms as stochastic and estimate (5) by random effects. However, consistency of the random effects estimator depends on the assumption that these terms are uncorrelated with the regressors (see Wooldridge, 2002, p. 257). Because the fixed-effects estimator is consistent whether this condition is satisfied or not, I treat the μ_i as a set of constants to be estimated.

Table 6

Inequality and the Variance of Computer Usage

Dependent variable	College fraction	Computer rate	Computer variance
Overall variance	0.08* (0.03)	0.05* (0.02)	0.2* (0.06)
Overall 90-10 difference	0.26* (0.1)	0.13* (0.07)	0.61* (0.18)
Overall 90-50 difference	–0.01 (0.06)	–0.03 (0.04)	0.37* (0.11)
Overall 50-10 difference	0.27* (0.06)	0.16* (0.04)	0.24* (0.12)
College–no high school	0.16 (0.16)	0.32* (0.12)	0.1 (0.3)
College–some high school	0.24* (0.13)	0.16* (0.08)	0.14 (0.16)
College–high school	0.13 (0.09)	0.01 (0.05)	0.48* (0.12)
College–some college	0.07 (0.07)	0.01 (0.04)	0.29* (0.09)
Residual variance	0.075* (0.03)	–0.0004 (0.02)	0.11* (0.04)
Residual 90-10 difference	0.2* (0.08)	0.03 (0.05)	0.32* (0.13)
Residual 90-50 difference	0.11* (0.05)	0.01 (0.04)	0.2* (0.09)
Residual 50-10 difference	0.1* (0.04)	0.02 (0.03)	0.12* (0.07)

NOTE: Coefficients from estimation of Specification III of (4) in which the variance of computer usage has also been added. Heteroskedasticity-consistent standard errors are reported in parentheses; * denotes significance at the 10 percent level.

Table 6 shows results from the inclusion of the variance of computer usage. For the sake of brevity, I have reported only the output from specification III, in which $X_{i,t}$ contains the college graduate, computer usage, and union membership fractions. While there is a slight dropoff in some of the magnitudes of the coefficients relative to the baseline estimates in Table 4, most are little changed after including the variance of computer usage. In fact, the same coefficients that are significant in Table 4 are significant here as well.

Computer use variance itself also enters positively and, for the most part, significantly, just as one would expect. Each of the 12 coefficients reported in the table is positive, 10 significantly so. In these data, then, both the first and second moments of the distribution of computer usage correlate directly with wage inequality.[25]

Table 7 reports the results when both time- and industry-specific fixed effects are included

in the regressions. Two features of the results are especially notable. First, the majority of the coefficients on both the college employment fraction and frequency of computer use are positive, while those on the unionization rate are negative, just as in the baseline results. However, second, the number of coefficients that differ statistically from zero at conventional levels has dropped relative to the results in Table 4. To be sure, among the college fraction coefficients, more than half (15 of 24) remain significant, indicating that industry-specific changes in many of the inequality measures correlate strongly with changes in their fractions of highly educated labor. At the same time, only 13 of the 36 unionization coefficients and 3 of the 24 computer usage coefficients differ significantly from zero after conditioning on time-invariant industry terms.

Very likely, this decrease in significance stems from the decline in the extent of variation in the data once industry-specific intercepts are included. This particular aspect of the estimation can be inferred from the sharp rise in the goodness-of-fit statistics reported in the final columns of Tables 4 and 7. The average R^2 rises from 0.21 to 0.67 with the addition of the industry effects.

---

[25] Since the majority of industry-year observations have computer usage fractions less than 0.5, it is not surprising that the mean and variance of the computer usage distribution are positively associated. However, the correlation is relatively modest, 0.47, suggesting that each variable may reasonably pick up some of the variation in inequality independently of the other.

Table 7

Inequality Regressions—Industry Effects Included

Dependent variable	Specification	College fraction	Computer rate	Union rate	R^2
Overall variance	I	0.25* (0.02)	—	–0.09* (0.02)	0.8
	II	—	0.08* (0.05)	–0.16* (0.09)	0.8
	III	0.24* (0.08)	0.07 (0.05)	–0.12 (0.09)	0.81
Overall 90-10 difference	I	0.68* (0.12)	—	–0.5* (0.09)	0.77
	II	—	0.08 (0.17)	–0.23 (0.4)	0.76
	III	1.07* (0.4)	0.03 (0.17)	–0.13 (0.4)	0.77
Overall 90-50 difference	I	0.17* (0.1)	—	–0.34* (0.07)	0.67
	II	—	–0.04 (0.11)	–0.03 (0.25)	0.7
	III	0.28 (0.4)	–0.06 (0.12)	0.02 (0.3)	0.7
Overall 50-10 difference	I	0.5* (0.06)	—	–0.16* (0.06)	0.69
	II	—	0.12 (0.12)	–0.2 (0.24)	0.72
	III	0.79* (0.16)	0.09 (0.12)	–0.16 (0.22)	0.74
College–no high school	I	0.08 (0.14)	—	–0.24* (0.11)	0.59
	II	—	0.06 (0.15)	–0.18 (0.27)	0.67
	III	0.32 (0.4)	0.03 (0.16)	–0.13 (0.3)	0.67
College–some high school	I	0.17 (0.11)	—	–0.18* (0.08)	0.6
	II	—	0.25* (0.12)	–0.32* (0.19)	0.67
	III	0.27 (0.3)	0.25* (0.12)	–0.29 (0.2)	0.67
College–high school	I	0.04 (0.07)	—	–0.16* (0.06)	0.6
	II	—	0.1 (0.08)	–0.15 (0.13)	0.7
	III	0.1 (0.2)	0.1 (0.08)	–0.12 (0.14)	0.7
College–some college	I	0.007 (0.07)	—	–0.06 (0.06)	0.44
	II	—	0.095 (0.08)	0.05 (0.14)	0.57
	III	–0.17 (0.16)	0.11 (0.08)	0.007 (0.14)	0.57
Residual variance	I	0.13* (0.02)	—	–0.04* (0.02)	0.74
	II	—	0.04 (0.03)	–0.13 (0.08)	0.7
	III	0.13* (0.07)	0.03 (0.03)	–0.16* (0.08)	0.7
Residual 90-10 difference	I	0.44* (0.1)	—	–0.16* (0.07)	0.7
	II	—	0.03 (0.14)	0.28 (0.3)	0.69
	III	0.78* (0.3)	–0.01 (0.1)	0.17 (0.3)	0.7
Residual 90-50 difference	I	0.21* (0.07)	—	–0.03 (0.05)	0.59
	II	—	–0.08 (0.11)	0.09 (0.22)	0.58
	III	0.43* (0.2)	–0.1 (0.1)	–0.02 (0.2)	0.59
Residual 50-10 difference	I	0.23* (0.06)	—	–0.13* (0.05)	0.53
	II	—	0.11 (0.07)	0.2 (0.17)	0.59
	III	0.35* (0.18)	0.09 (0.07)	0.19 (0.16)	0.59

NOTE: Coefficients from estimation of (5). Specification I uses annual data 1983-2002. Specifications II and III use only data from 1984, 1989, 1993, and 1997. Each regression includes year dummies and industry-specific fixed effects. Heteroskedasticity-consistent standard errors are reported in parentheses; * denotes significance at the 10 percent level.

While this necessarily tempers the conclusions that can be drawn from the results, one can still interpret this evidence as consistent with the SBTC hypothesis.

CONCLUSIONS

Despite the presence of a large literature examining the rise of earnings inequality in the United States, surprisingly few studies have directly explored the role of information technology in driving this trend. Such an omission is particularly surprising in light of the general consensus that has emerged in support of the skill-biased technological change hypothesis. This paper has attempted to offer some evidence on this issue.

The findings indicate that the vast majority of the rise in U.S. wage inequality over the past two decades is the product of increasing gaps between workers within the same industry rather than between workers across different industries. This result holds whether considering workers of differing levels of observable skill (overall inequality) or those with the same levels (residual inequality). What is more, within-industry inequality—defined in overall, residual, and between-education-group terms—tends to be positively associated with the two measures of skill-biased technological change considered here, the college employment fraction and the frequency of computer usage. Collectively, these two observations are compatible with the idea that skill-biased technological change has been a significant element in the rise of wage dispersion in the United States.

Of course, because the two measures of skill-biased technological change considered here are less than ideal, there remains ample room for additional research on this topic. In particular, studies examining the extent to which plants and industries have adopted specific production technologies, such as those considered by Dunne (1994), and how the implementation of those technologies correlate with earnings differentials would greatly assist in clarifying the skill-biased technological change hypothesis. Considering the popularity of the theory, such an undertaking certainly seems worthwhile.

REFERENCES

Acemoglu, Daron. "Changes in Unemployment and Wage Inequality: An Alternative Theory and Some Evidence." *American Economic Review*, December 1999, *89*(5), pp. 1259-78.

Acemoglu, Daron. "Technical Change, Inequality, and the Labor Market." *Journal of Economic Literature*, March 2002, *40*(1), pp. 7-72.

Autor, David H.; Katz, Lawrence F. and Krueger, Alan B. "Computing Inequality: Have Computers Changed the Labor Market?" *Quarterly Journal of Economics*, November 1998, *113*(4), pp. 1169-213.

Autor, David H.; Levy, Frank and Murnane, Richard J. "The Skill Content of Recent Technological Change: An Empirical Exploration." *Quarterly Journal of Economics*, November 2003, *118*(4), pp. 1279-333.

Berman, Eli; Bound, John and Griliches, Zvi. "Changes in the Demand for Skilled Labor within U.S. Manufacturing: Evidence from the Annual Survey of Manufactures." *Quarterly Journal of Economics*, May 1994, *109*(2), pp. 367-97.

Blau, Francine D. and Kahn, Lawrence M. "International Differences in Male Wage Inequality: Institutions versus Market Forces." *Journal of Political Economy*, August 1996, *104*(4), pp. 791-837.

Borjas, George J.; Freeman, Richard B. and Katz, Lawrence F. "How Much Do Immigration and Trade Affect Labor Market Outcomes?" *Brookings Papers on Economic Activity*, 1997, (1), pp. 1-90.

Card, David. "The Causal Effect of Education on Earnings," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics.* Volume 3A. Amsterdam: Elsevier, 1999, pp. 1801-63.

Card, David and DiNardo, John E. "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles." *Journal of Labor Economics*, October 2002, *20*(4), pp. 733-83.

Caselli, Francesco. "Technological Revolutions." *American Economic Review*, March 1999, *89*(1), pp. 78-102.

Doms, Mark; Dunne, Timothy and Troske, Kenneth R. "Workers, Wages, and Technology." *Quarterly Journal of Economics*, February 1997, *112*(1), pp. 253-90.

Dunne, Timothy. "Plant Age and Technology Use in U.S. Manufacturing Industries." *RAND Journal of Economics*, Autumn 1994, *25*(3), pp. 488-99.

Dunne, Timothy; Foster, Lucia; Haltiwanger, John and Troske, Kenneth R. "Wage and Productivity Dispersion in United States Manufacturing: The Role of Computer Investment." *Journal of Labor Economics*, April 2004, *22*(2), pp. 397-429.

Fortin, Nicole M. and Lemieux, Thomas. "Institutional Changes and Rising Wage Inequality: Is There a Linkage?" *Journal of Economic Perspectives*, Spring 1997, *11*(2), pp. 75-96.

Johnson, George E. "Changes in Earnings Inequality: The Role of Demand Shifts." *Journal of Economic Perspectives*, Spring 1997, *11*(2), pp. 41-54.

Juhn, Chinhui; Murphy, Kevin M. and Pierce, Brooks. "Wage Inequality and the Rise in Returns to Skill." *Journal of Political Economy*, June 1993, *101*(3), pp. 410-42.

Katz, Lawrence F. and Murphy, Kevin M. "Changes in Relative Wages, 1963-1987: Supply and Demand Factors." *Quarterly Journal of Economics*, February 1992, *107*(1), pp. 35-78.

Krueger, Alan B. "How Computers Have Changed the Wage Structure: Evidence from Microdata, 1984-1989." *Quarterly Journal of Economics*, February 1993, *108*(1), pp. 33-60.

Levy, Frank and Murnane, Richard J. "U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations." *Journal of Economic Literature*, September 1992, *30*(3), pp. 1333-81.

Park, Jim Heum. "Estimation of Sheepskin Effects and Returns to Schooling Using the Old and the New CPS Measures of Educational Attainment." Working Paper No. 338, Industrial Relations Section, Princeton University, December 1994.

Topel, Robert H. "Factor Proportions and Relative Wages: The Supply-Side Determinants of Wage Inequality." *Journal of Economic Perspectives*, Spring 1997, *11*(2), pp. 55-74.

Wheeler, Christopher H. "Cities, Skills, and Inequality." Working Paper No. 2004-020A, Federal Reserve Bank of St. Louis, September 2004.

Wooldridge, Jeffrey M. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press, 2002.

APPENDIX

Data

Hourly wages are calculated as the ratio of a worker's weekly earnings to usual hours worked per week. The MORG files do topcode weekly earnings at various points ($999 for 1983 to 1988, $1,923 for 1989 to 1997, and $2,884 for 1998 to 2002). For topcoded values, I follow Card and DiNardo (2002) and impute the weekly wages as 1.4 times the topcode value to approximate the mean of the upper tail of the wage distribution. Similar techniques have been used by Katz and Murphy (1992), Juhn, Murphy, and Pierce (1993), and Autor, Katz, and Krueger (1998). All hourly wages are converted to real terms ($2,000) using the personal consumption expenditure chain-type price index. Once these values are computed, I then restrict the sample to workers with hourly wage earnings between $2.60 (which is slightly in excess of one-half the current federal minimum wage) and $150 to eliminate outliers. Inequality calculations are computed using the CPS "earnings" weight. For the educational attainment and union

membership rates, a total of 2,693,370 observations are used. This corresponds to an average of 604.1 observations per industry-year (minimum = 2, maximum = 9,672). For the inequality calculations, a total of 1,156,715 observations are used. This corresponds to 258.8 observations per industry-year (minimum = 2, maximum = 7,090).

Since educational attainment is not coded in the CPS as years of schooling completed for all years (i.e., there was a change in the education variable between 1991 and 1992), I follow previous work (e.g., Autor, Katz, and Krueger, 1998) and impute values from Table 5 of Park (1994), which establishes a correspondence between the old and new CPS education variables. Potential work experience is then calculated as the maximum of 0 and (age-years of education − 6).

Computer usage by industry is calculated using responses to the question: "Do you directly use a computer at work?" reported in the October supplements for the years 1984, 1989, 1993, and 1997. Here, I use the CPS "supplement" weights for 1984, 1989, and 1993 and the "final" weight for 1997. In all, there are 59,642 observations for 1984, 60,304 for 1989, 54,273 for 1993, and 50,478 for 1997. The mean number of observations per industry-year is 308.2 (minimum = 7, maximum = 4,298).

A consistent set of 230 industries (ranging in number from 219 to 230 in any given year) are identified over the 20-year period. Because the CPS industry codes changed in 1992, a consistent classification scheme was implemented using the crosswalk provided by the U.S. Bureau of the Census (and summarized by Barry Hirsch at his website: www.trinity.edu/bhirsch).

Derivation of Variance Expression

To show that equations (1) and (3) give equivalent expressions for the variance of wages, note first that (1) expands to

(1′)
$$V_t = \frac{1}{N_t}\sum_{i=1}^{I_t}\sum_{j=1}^{N_{i,t}}w_{j,i,t}^2 - \frac{2}{N_t}\sum_{i=1}^{I_t}\sum_{j=1}^{N_{i,t}}w_{j,i,t}\bar{w}_t + \frac{1}{N_t}\sum_{i=1}^{I_t}\sum_{j=1}^{N_{i,t}}\bar{w}_t^2,$$

whereas (3) can be written as

(3′)
$$V_t = \frac{1}{N_t}\sum_{i=1}^{I_t}\sum_{j=1}^{N_{i,t}}w_{j,i,t}^2 - \frac{2}{N_t}\sum_{i=1}^{I_t}\sum_{j=1}^{N_{i,t}}w_{j,i,t}\bar{w}_{i,t} + \frac{2}{N_t}\sum_{i=1}^{I_t}\sum_{j=1}^{N_{i,t}}\bar{w}_{i,t}^2 - \frac{2}{N_t}\sum_{i=1}^{I_t}\sum_{j=1}^{N_{i,t}}\bar{w}_t\bar{w}_{i,t} + \frac{1}{N_t}\sum_{i=1}^{I_t}\sum_{j=1}^{N_{i,t}}\bar{w}_t^2.$$

Because

$$\frac{2}{N_t}\sum_{i=1}^{I_t}\sum_{j=1}^{N_{i,t}}w_{j,i,t}\bar{w}_{i,t} = \frac{2}{N_t}\sum_{i=1}^{I_t}\bar{w}_{i,t}\bar{w}_{i,t}N_{i,t} = \frac{2}{N_t}\sum_{i=1}^{I_t}\bar{w}_{i,t}^2N_{i,t}$$

and

$$\frac{2}{N_t}\sum_{i=1}^{I_t}\sum_{j=1}^{N_{i,t}}\bar{w}_{i,t}^2 = \frac{2}{N_t}\sum_{i=1}^{I_t}\bar{w}_{i,t}^2N_{i,t},$$

the second and third terms on the right-hand side of (3′) sum to zero, leaving

(3″)
$$V_t = \frac{1}{N_t}\sum_{i=1}^{I_t}\sum_{j=1}^{N_{i,t}}w_{j,i,t}^2 - \frac{2}{N_t}\sum_{i=1}^{I_t}\sum_{j=1}^{N_{i,t}}\bar{w}_t\bar{w}_{i,t} + \frac{1}{N_t}\sum_{i=1}^{I_t}\sum_{j=1}^{N_{i,t}}\bar{w}_t^2.$$

Given that the middle term on the right-hand side of (3″) can be expressed as

$$\frac{2\bar{w}_t}{N_t}\sum_{i=1}^{I_t}\bar{w}_{i,t}N_{i,t} = \frac{2\bar{w}_t}{N_t}\sum_{i=1}^{I_t}N_{i,t}\frac{1}{N_{i,t}}\sum_{j=1}^{N_{i,t}}w_{j,i,t} = \frac{2\bar{w}_t}{N_t}\sum_{i=1}^{I_t}\sum_{j=1}^{N_{i,t}}w_{j,i,t} = \frac{2}{N_t}\sum_{i=1}^{I_t}\sum_{j=1}^{N_{i,t}}w_{j,i,t}\bar{w}_t,$$

equations (1′) and (3′) are equivalent. Therefore, (1) and (3) are equivalent.

Table A1

Summary Statistics

Variable	Mean	Standard deviation	Minimum	Maximum	Observations
Overall variance	0.25	0.09	0	0.97	4,381
Overall 90-10 difference	1.26	0.27	0	2.82	4,381
Overall 90-50 difference	0.67	0.2	0	2.64	4,381
Overall 50-10 difference	0.59	0.17	0	1.76	4,381
College–no high school	0.68	0.35	–1.36	2.05	3,521
College–some high school	0.61	0.31	–1.17	2.02	3,943
College–high school	0.45	0.26	–0.96	2.14	4,251
College–some college	0.34	0.25	–0.81	2.51	4,263
Residual variance	0.18	0.07	0	0.67	4,379
Residual 90-10 difference	1.03	0.22	0	2.53	4,379
Residual 90-50 difference	0.52	0.15	0	1.82	4,379
Residual 50-10 difference	0.51	0.14	0	1.86	4,379
College fraction	0.21	0.14	0	0.75	4,384
Computer usage rate	0.38	0.23	0	1	887
Union membership rate	0.13	0.13	0	0.84	4,384

NOTE: Summary statistics for selected industry characteristics over the period 1983-2002.