THE NEW WORKER-EMPLOYER CHARACTERISTICS DATABASE[1]

Kimberly Bayard, U.S. Census Bureau; Judith Hellerstein, University of Maryland and NBER; David Neumark, Michigan State University and NBER; Kenneth R. Troske, University of Missouri
Kimberly Bayard, U.S. Census Bureau, Center for Economic Studies, 4700 Silver Hill Road, Stop 6300, Washington, DC 20233-6300 USA kbayard@ces.census.gov

ABSTRACT

The New Worker-Employer Characteristics Database (NWECD) is a matched employer-employee data set that includes longitudinal information for both manufacturing and non-manufacturing establishments. Individual and household data come from the 1990 Decennial Census of Population Sample Edited Detail File. The information on establishments comes from the Census Bureau's Standard Statistical Establishment List (SSEL). Unlike most other matched worker-employer data sets for the U.S., the NWECD covers all regions of the country and all sectors of the economy. In addition, the rich demographic information on workers make it a useful source for determining workforce characteristics of different types of establishments. This paper introduces the NWECD, documents its construction, and evaluates its representativeness. Although the NWECD is not representative of the underlying populations, regressions using the NWECD produce coefficient estimates similar to those in the underlying representative data sets. The NWECD is therefore useful for conducting empirical research that requires the use of matched employer-employee data.

Key Words: Matched Employer-Employee Data, Data Set Construction, Data Set Evaluation

1. INTRODUCTION

One of the fundamental issues in empirical labor economics is understanding the determinants of pay. Most models of wage determination are based on actions or characteristics of both employers and employees, and the testing of these models requires data that are sufficiently rich in information about both parties. Unfortunately, for the U.S. economy, there exist very few data sets that link workers to employers and therefore contain the detail necessary to fully address questions regarding wage determination, as well as other topics in labor economics.

In this paper, we present the New Worker-Establishment Characteristics Database (NWECD), a new data set that matches workers to their employers. The NWECD is a cross-sectional data set that contains information collected by the U.S. Census Bureau from individuals and their employers in 1990. The data set includes observations on over one million individuals matched to over 150,000 establishments, and covers all sectors of the economy and all regions of the country. The NWECD is an extension of the Worker-Establishment Characteristics Database (WECD), a database that contains information on manufacturing workers and their employers.[2] In recent work, we have used the NWECD to explore the effects of workplace segregation on gender, race, and ethnic wage gaps (Bayard et al. 1998a and 1998b).

Both the NWECD and the WECD are created from two underlying data sources: the Sample Edited Detail File (SEDF), which contains all individual responses to the 1990 Decennial Census one-in-six long-form, and the 1990 Standard Statistical Establishment List (SSEL), an administrative database that contains information for all establishments operating in the United States in 1990. Following the procedures used to create the WECD, we use detailed location and industry information available on the SEDF and the SSEL to identify establishments where workers are employed.

2. THE MATCHING PROCESS

The NWECD is created by matching together worker records from the SEDF with employer records from the SSEL. The SEDF contains information about the detailed industry and location of respondents' employers, and the SSEL contains similar information for each establishment. To create the NWECD, we link workers with employers that have identical industry and location information.

---

[1]This paper reports the results of research and analysis undertaken by Census Bureau staff and co-authors. It has undergone a more limited review by the Census Bureau than its official publications. This report is released to inform interested parties and to encourage discussion.

[2]The construction and evaluation of the WECD is well documented in Troske (1998).

The NWECD is created by matching workers who responded to the long-form of the 1990 Decennial Census to their employers. This match is possible because the U.S. Census Bureau maintains a register of all businesses operating in the United States. To summarize the matching process, we identify the establishments where workers are employed by linking together detailed industry and location information available on both the SEDF and the SSEL.

The 1990 Decennial Census long-form asked households to report the name and address of the employer in the previous week for each employed member of the household. The long-form also asked respondents to name and describe in a word or two the type of business or industry of the most recent employer of all members of the household. Consistent with coding procedures for household surveys, the Census Bureau uses the respondent-provided information to assign geographic and industry codes to each record in the data. In addition to these codes, the SEDF contains standard demographic characteristics. The NWECD draws data from the records of more than 17 million respondents who indicated they were employed in week prior to receiving a Census form.

The SSEL is an annual list of all business establishments in the U.S. that the Census Bureau uses as a sampling frame for Economic Censuses and Surveys. The SSEL contains each establishment's name and address, a geographic code based on location, and a four-digit Standard Industrial Classification (SIC) code. The SSEL also contains information on the number of employees and total annual payroll. There are unique establishment identifiers that allow establishments to be linked to other establishments that are part of the same enterprise. To construct the NWECD, we selected from the 1990 SSEL the 5.6 million private sector records.[3]

Matching workers to employers involves four steps. The first is standardizing geographic and industry codes in the two data sets. Next, we selected all establishments unique in an industry-location cell. All workers who reported working in the same industry-location cell as a unique establishment were then assigned to the establishment. Finally, we eliminated all matches based on imputed data.

A number of issues in the matching process merit further discussion. One is standardizing the geographic and industry codes. The Census Bureau divides the country into a hierarchy of geographic areas. For our purposes, the relevant areas are state, county, place, tract, and block. The Census Bureau assigns a unique code to every state in the country, and within each state, to every county. Population centers with 2,500 or more people are assigned unique place codes. Because these population centers are unique within a state, but can cross county boundaries, we distinguish between areas in the same place located in different counties. The Census Bureau also divides populated counties into distinct tracts and blocks.[4] Thus, an establishment located in a metropolitan area has a unique geographic code that identifies the establishment's state, county, place, tract, and block.

Another coding issue is that the Census Bureau codes the SSEL slightly differently than the SEDF. Whereas the SEDF has state, county, place, tract, and block codes for all records, the SSEL has tract and block codes for only a subset of its records. Before 1992, the SSEL contained no tract and block codes.[5] To assign tract and block codes to records in the 1990 SSEL, we matched establishment records from the 1990 SSEL to their 1992 counterparts, and assigned the available tract and block codes from the 1992 SSEL to the analogous 1990 establishment records. For establishments not in the 1992 SSEL or those that lacked tract and block information we assigned missing values.

---

[3]We exclude establishments in Public Administration in the discussion below due to a consolidation of establishment records in some jurisdictions in the SSEL. Because government establishment records are occasionally merged into one entry in the SSEL, it is impossible to ensure unique worker-establishment links in this sector. To avoid incorrect matches, we exclude all workers and establishments in Public Administration.

[4]In some geographic areas, the Census Bureau uses Block Numbering Areas (BNAs) instead of tracts. For our purposes, a BNA is equivalent to a tract. The Census Bureau assigns tracts and blocks in tandem, so whenever an establishment is assigned a tract code, it is also always assigned a block code.

[5]One reason for limited tract and block coverage in the SSEL is incomplete address information. For some establishments, the only address information is the mailing address of the business, and not the physical address. This mailing address may be a P.O. box, which cannot be assigned a tract or block code. In 1992, the Census Bureau assigned tract and block codes to 45 percent of the records in the SSEL.

To build the NWECD, we also standardize industry codes across the SSEL and the SEDF. The industry code in the SSEL is SIC-based whereas the industry code on the SEDF comes from the Census Industry Classification (CIC) system. Because CIC codes are more aggregate, we use a concordance table to convert SIC codes to CIC codes.[6]

The next step in matching workers to employers is to keep only those establishments that are unique in an industry-location cell. Recall that for all establishments in the SSEL we have state, county, and place codes, while for a subset of establishments we also have tract and block codes. In order to select establishments that have unique industry-location information we first keep establishments that are unique within an industry-state-county-place cell. The remaining establishments are not unique at the place level, but may be unique at the tract or block level. In order to determine which establishments are unique at the tract or block level, we first keep only those establishments in an industry-place cell where all establishments in this cell have non-missing tract and block codes. We then keep those establishments that are unique in an industry-state-county-place-tract-block cell. This produces a data set with 385,135 establishments available for matching. We then assign workers to industry-location cells, and match all workers who are in the same industry-location cell to the corresponding establishment.

To ensure that workers are correctly matched with employers, we discard matches based on imputed data. Data are imputed for several reasons. For example, some long-form respondents fail to provide sufficient address or industry information for their employers. When this occurs, the workers' geographic or industry information is imputed. Alternatively, an establishment's record in the SSEL may have an incomplete SIC code. In these cases, Census randomly assigns the additional digits necessary to complete it.

We also discard matches when the number of individuals matched to an establishment exceeds the number of employees that the establishment reports in the SSEL. One possible reason for this "over matching" is the nature of the place-of-work question on the long-form. Because the long-form asks respondents for the location where they worked in the previous week, individuals who worked at a site other than their primary employer's location may be improperly assigned to an establishment. A second reason the number of individuals matched might exceed an establishment's total employment is the time lag between when the Census Bureau surveys workers and employers. Employers are asked to report employment on March 12th, and workers are asked to report where they worked on (roughly) April 1st. If total employment is higher on April 1st than on March 12th, the number of workers matched to the establishment may exceed total employment. Another problem is that total employment in the SSEL covers only employees whereas both workers and owners are accounted for in the SEDF and are matched to establishments.

Despite these benign reasons why the number of matched workers may exceed reported establishment employment, cases of "over matching" may reflect errors in industry or location coding in either the SEDF or the SSEL. To avoid potentially incorrect matches, we discard cases where this occurs.[7] The resulting data set contains 1,056,635 workers matched to 153,291 establishments.

3. EXAMINING THE REPRESENTATIVENESS OF THE NWECD

The two predominant features of the NWECD are that it contains information on the workforce composition of establishments, and also on the characteristics of establishments where workers are employed. From a research perspective, one would like to know the degree to which the NWECD is representative of the underlying populations of workers and establishments. To assess the representativeness of the NWECD, we compare descriptive statistics for workers and establishments in the NWECD with those in the SSEL and SEDF.

Table 1 presents comparisons of characteristics for the three levels of establishment data: the full SSEL, the set of establishments unique in an industry-location cell, and the NWECD. As mentioned earlier, we only attempt to match

---

[6]In general, CIC codes correspond to three-digit SIC codes. One exception is construction where one CIC code corresponds to three two-digit SIC codes. In addition, some SIC industries map to more than one CIC code. We omit the few establishments in these industries.

[7]In addition, we eliminate workers with missing or zero reported earnings or who work outside the U.S.

workers to those establishments unique in an industry-location cell. NWECD establishments are those "unique" establishments that we successfully match to workers. In Panel A, we display the number of establishments and the average employment for all establishments in the data. Subsequent panels break down the overall totals by whether or not the establishment is located in an MSA (Panel B), by size (Panel C), and by industry (Panel D). Columns (1)-(3) present the number of SSEL, unique, and NWECD establishments, respectively. Column (4) presents the proportion of SSEL establishments unique in an industry-location cell, while column (5) presents the proportion of SSEL establishments in the NWECD. Columns (6)-(8) present mean employment for each group of establishments.

Panel A shows the effect of the matching strategy on the overall sample of establishments. Of the 5.6 million establishments in the SSEL, only 6.9 percent or 385,135 can be assigned to a unique industry-location cell. In addition, these unique establishments are almost 80 percent larger than the typical establishment, averaging nearly 36 employees compared with 20 employees in the typical SSEL establishment. This row also shows that uniqueness does not guarantee that an establishment is matched, as only 40 percent of unique establishments are in the NWECD. There are 153,291 establishments in the NWECD, representing 2.7 percent of SSEL establishments. Matched establishments, averaging 72 employees, tend to be even larger than unique establishments. This increase in average size is the result of two factors. First, because the long-form was sent to only 1-in-6 households, large establishments are more likely to employ workers receiving a long-form; and second, smaller establishments are more likely to be eliminated from the data because of imputed data items or because the number of matched workers exceeds reported employment.[8]

The numbers in Panels B and C show that the probability of being unique and the probability of appearing in the NWECD vary systematically with the location and size of the establishment. Panel B shows that establishments located outside of an MSA are more than twice as likely to be located in a unique industry-location cell and to appear in the NWECD. Columns (6)-(8) show that, for establishments both within and outside of an MSA, the matching strategy produces a data set with establishments that are substantially larger on average than establishments in the SSEL. The Panel C results show that the probability that an establishment is unique in an industry-location cell, and the probability that an establishment appears in the NWECD, increases monotonically with the size of the establishment. Only 6.4 percent of establishments with fewer than 10 employees are located in a unique industry-location cell and only 1.8 percent of these establishments appear in the NWECD. In contrast, 20 percent of establishments with 500 or more employees are in unique industry-location cells and 19 percent of these establishments appear in the NWECD.

Panel D of Table 1 shows that the match rate varies substantially by industry. Almost 23 percent of all Manufacturing establishments, but less than one percent of those in Construction appear in the NWECD. In part, this variation results from finer definition of industry in some sectors than others. For example, under the Census Industrial Classification System, Manufacturing is divided into 82 sub-categories, whereas all of Construction is consolidated into one. Consequently, for any given location, it is more likely that a Manufacturing establishment will be located in a unique industry-location cell, and subsequently matched to workers, than an establishment in Construction.

An additional explanation for the disparity in match rates across industries relates to differences in average establishment size. In general, industries with larger average establishment size are more likely to be located in unique industry-location cells, and also more likely to appear in the NWECD. Because the long-form was sent to only one-in-six households, larger establishments are more likely to employ at least one worker who received the form, and therefore more likely to appear in the matched data set. Consequently, industries with larger establishments appear in the NWECD with greater frequency than those with small establishments. We see evidence of this in Panel D of Table 1. The largest establishments on average are in Manufacturing, and 12 percent of all establishment in this sector appear in the NWECD. In contrast, Retail and Wholesale establishments are much smaller, and only two to three percent of all establishments in these sectors appear in the matched data set.

Table 2 compares the number and annual earnings of workers in the SEDF with workers in the NWECD for all workers (Panel A), by whether or not a worker's employer is located in an MSA (Panel B) and by one-digit industry (Panel C).

[8]Smaller establishments have less reliable industry information than large establishments because they are more likely to be recently formed, and less likely to be covered by one of the Economic Censuses. Consequently, they are more likely to have their industry code imputed on the SSEL.

In all comparisons, we restrict attention to individuals with positive annual earnings. Columns (1) and (2) present the number of workers in the SEDF and NWECD, respectively. Column (3) presents the proportion of workers matched to an establishment (column (2)/column (1)). Columns (4) and (5) present the mean of worker earnings in the SEDF and NWECD, while column (6) presents the log difference in the average worker earnings estimates.

Panel A shows that there are more than 14 million workers in the SEDF. Of these, we match just over one million, or 7.4 percent, to their employers. Recall from Table 1 that approximately 2.7 percent of all SSEL establishments appear in the NWECD. The higher match rate for workers than establishments is not surprising because workers are more likely to be employed in large establishments which are more likely to be in unique industry-location cells. Columns (4)-(6) show that average earnings of NWECD workers are very close to average earnings of SEDF workers.

The numbers in Panels B and C show that the distribution of matched workers across location and industry is similar to the distribution of matched establishments. Panel B shows that we match 12.5 percent of all SEDF individuals who work outside of an MSA and only 5.7 percent of workers employed in urban counties. Panel C indicates a higher proportion of workers matched in Manufacturing, Services, and Transportation than in Construction, FIRE, Wholesale, or Retail. Columns (4)-(6) demonstrate that average annual earnings of SEDF and NWECD workers vary by location and industry. These columns also show that earnings of SEDF and NWECD workers are closer in some industries than others. For instance, the earnings of NWECD workers in Manufacturing, Transportation, and Retail are quite similar to SEDF averages, while those of workers in FIRE, Construction, Wholesale, and Services are fairly different.

Table 3 continues the comparisons of characteristics of workers in the NWECD and SEDF. Columns (1) and (2) compare characteristics for all workers, columns (3) and (4) compare workers earning between $2.50 and $500 per hour, and columns (5) and (6) cover workers in this wage range who also report usually working over 30 hours per week and at least 30 weeks in 1989. We define this last set of workers as "full-time". The numbers in Table 3 point to some differences between the workers represented in the NWECD and those in the entire SEDF. NWECD workers are slightly less likely to be black or Hispanic, and more likely to be female, married, or full-time workers than individuals in the SEDF. In addition, NWECD workers are more likely laborers and to work in manufacturing and services. NWECD workers are also slightly older, and are more likely to have a high school degree but less likely to have no high school education or to have a bachelor's or advanced degree. NWECD workers tend to work more weeks in the previous year, but have slightly lower earnings and hourly wages.

The results in Tables 1-3 raise concerns about the representativeness of the NWECD. Establishments in the NWECD are substantially larger than those in the SSEL. In addition, establishments and workers in Manufacturing, and those located outside of an MSA are more likely to appear in the NWECD than their non-Manufacturing and urban counterparts. The demographic characteristics of workers also differ somewhat across the two data sets. From these differences, it is clear that the NWECD is not a random sample of the SEDF.

The overall effects of this non-random matching depend on the questions being addressed with these data. In particular, the non-representative nature of the NWECD may render it of little value in constructing population estimates, but may have little impact on estimated conditional correlations (or regression relationships). To further gauge the usefulness of these data, we examine whether the NWECD can replicate relationships found in the SEDF between workers' characteristics and their wages and hours worked. Columns (1)-(3) in Table 4 present coefficient estimates and standard errors from regressions of the logs of hourly worker wages on a standard set of characteristics, and columns (4)-(6) present analogous results for regressions using the log of annual hours as a dependent variable. The first column of each group presents basic results for the SEDF, the next column adds a control for whether the worker is matched to an establishment, and the last repeats the wage regression of the first column, but uses the NWECD. The coefficients in the wage regressions reported in columns (1)-(3) are very similar across the data sets. In both samples, female workers earn 10-11 percent lower wages. Black and Hispanic workers also face similar wage penalties in each sample, and married workers earn comparable premiums (19-20 percent). The relationship between education and wages is also similar across data sets, as are the penalties faced by black and Hispanic women. The coefficient on the match variable in column (2) shows that, controlling for standard worker characteristics, matched workers earn 2.6 percent lower wages.

The results in columns (4)-(6) of Table 4 list coefficients and standard errors from regressions of the log of annual hours on demographic characteristics. In general, the NWECD and SEDF results are similar, although they are not as close

as the wage regression results. In the NWECD, women work 7.7 percent fewer hours annually than men, and in the SEDF, 9 percent. The relationship between age and annual hours worked is also close in both samples. In both data sets, blacks work approximately 6 percent fewer hours than whites, and married workers work 12-13 percent more hours per year. One difference in the samples is the disparity in hours worked between more and less educated workers. For example, in the NWECD, workers with a high school degree work 12 percent more hours annually than workers without a degree, while in the SEDF, the difference is 17 percent. The margins of difference between more and less educated groups are consistent across both samples which suggests that it is the least educated workers in the SEDF who work fewer hours than their counterparts in the NWECD.

The analysis of representativeness of the NWECD suggests that it is not a representative sample of the underlying population of establishments or workers. However, it appears that the non-representativeness is unlikely to introduce much bias into regression estimates of general interest, especially hourly wage regression estimates.

4. CONCLUSION

This paper introduces the New Worker-Establishment Characteristics Database, a data set that matches workers to the establishments where they work. Despite the research benefits of having joint information on workers and their employers, there are very few matched data sets for the U.S. Of the matched data sets that do exist, the NWECD has strengths that make it uniquely suited to address a number of interesting research questions. The NWECD covers workers and establishments in all sectors of the economy and all regions of the country. It contains observations for multiple workers at the same establishment, thereby providing a profile of an establishment's workforce. The data set also contains a rich set of characteristics for each worker. These characteristics include gender, age, race, ethnicity, education, veteran's and citizenship status, and English proficiency. Because the Decennial Census is household-based, the NWECD can also be supplemented with detailed information about workers' households and neighborhoods.

This paper also documents the construction of the NWECD and analyzes its representativeness. We describe the steps we took to ensure accurate matches and thorough coverage of workers and establishments. Although the NWECD is not representative of the underlying populations, wage regressions on the NWECD produce coefficient estimates similar to those on the SEDF. The similarity in regression coefficients, especially those from hourly wage regressions, suggests that despite the unrepresentativeness of the NWECD, the data are useful for estimating many relationships that are of interest to economists.

5. REFERENCES

Bayard, K., J. Hellerstein, D. Neumark, and K. Troske (1998a), "New Evidence on Sex Segregation and Sex Differences in Wages from Matched Employee-Employer Data," NBER Working Paper No. 7003.

Bayard, K., J. Hellerstein, D. Neumark, and K. Troske (1998b),"Why are Racial and Ethnic Wage Gaps Larger for Men than for Women?" in J. Haltiwanger, J. Lane, J. Spletzer, J. Theeuwes, and K. Troske (eds.) *The Creation and Analysis of Employer-Employee Matched Data.* Amsterdam: Elsevier Science B.V., pp. 175-204.

Troske, K. (1998), "The Worker-Establishment Characteristics Database," in J. Haltiwanger, M. Manser, and R. Topel, (eds.) *Labor Statistics Measurement Issues.* Chicago: University of Chicago Press, pp. 371-403.

Table 1: Number, Proportion and Average Total Employment of SSEL, Unique, and Matched Establishments By Employment Size, Industry, and MSA status

	SSEL estab. (1)	Unique estab. (2)	NWECD estab. (3)	Proportion unique (4)	Proportion matched (5)	SSEL estab. empl. (6)	Unique estab. empl. (7)	NWECD estab. empl. (8)
				A. Total				
	5,587,650	385,135	153,291	0.069	0.027	20.11	35.82	72.39
				B. Location				
MSA	4,492,867	239,020	92,701	0.053	0.021	21.35	40.56	83.74
Non-MSA	1,091,700	146,085	60,590	0.134	0.051	14.94	28.06	55.03
				C. Establishment Size (Total Employment)				
1-9	3,955,604	255,041	72,123	0.064	0.018	3.57	3.62	4.34
10-24	943,383	64,210	32,485	0.068	0.034	14.91	15.01	15.35
25-49	351,123	25,806	16,465	0.073	0.047	34.28	34.55	34.85
50-99	182,558	17,366	12,814	0.095	0.070	68.72	70.18	70.62
100-249	106,274	13,794	11,435	0.130	0.108	150.35	152.98	154.42
250-499	28,807	4,887	4,293	0.170	0.149	342.67	348.53	350.03
500+	19,901	4,031	3,676	0.203	0.185	1,696.61	1,484.68	1,506.65
				D. Industry				
Agriculture	84,084	13,227	4,471	0.157	0.053	39.33	9.40	15.61
Mining	26,923	3,507	1,556	0.130	0.058	27.07	23.09	39.29
Construction	460,300	648	151	0.001	0.0003	11.14	6.11	12.03
Manufacturing	339,039	77,456	40,305	0.228	0.119	56.98	70.63	115.42
Transportation	206,078	28,839	14,529	0.140	0.071	34.28	60.58	107.35
Wholesale	427,506	41,098	13,370	0.096	0.031	15.92	11.40	18.20
Retail	1,329,908	83,592	30,127	0.063	0.023	14.76	11.04	18.83
FIRE	484,119	14,471	5,327	0.030	0.011	19.54	11.35	17.95
Services	1,779,285	122,297	43,455	0.069	0.024	19.63	39.35	88.53

Note: There are 450,408 SSEL establishments that have missing or non-classifiable SIC codes. There are 3,086 establishments in the SSEL and 30 establishments in unique industry place cells that we could not assign to an MSA because of incomplete geographic information in the SSEL.

Table 2: Number and Mean Earnings of SEDF and NWECD Workers By Industry and Location

	Number of SEDF workers (1)	Number of NWECD workers (2)	Proportion matched (3)	Mean earnings SEDF workers (4)	Mean earnings NWECD workers (5)	Log difference (6)
			A. Total			
	14,264,082	1,056,635	0.074	23,147.38 (7.749)	22,438.53 (22.031)	0.031
			B. Location			
MSA	10,751,733	616,994	0.057	24,932.18 (9.633)	24,692.94 (31.789)	0.010
Non-MSA	3,512,349	439,641	0.125	17,683.87 (10.461)	19,274.68 (27.826)	-0.086
			C. Industry			
Agriculture	333,628	12,002	0.036	16,069.73 (42.599)	16,966.45 (211.042)	-0.054
Mining	114,367	7,374	0.064	32,137.77 (86.359)	29,991.61 (260.715)	0.069
Construction	879,065	477	0.001	25,102.05 (29.596)	19,711.87 (1,003.29)	0.242
Manufacturing	2,933,974	441,810	0.151	26,730.98 (15.880)	25,468.95 (34.020)	0.048
Transportation	1,095,901	71,909	0.066	28,508.97 (22.747)	28,564.70 (82.533)	-0.002
Wholesale	668,366	30,721	0.046	28,277.24 (41.729)	23,718.03 (160.149)	0.176
Retail	2,471,348	88,067	0.036	14,837.84 (13.580)	14,205.60 (67.367)	0.044
FIRE	1,001,985	14,491	0.014	29,094.73 (43.285)	20,979.85 (204.333)	0.327
Services	4,765,448	389,784	0.082	21,966.55 (13.814)	19,714.97 (34.966)	0.108

Note: The numbers in parentheses are standard errors of means.

Table 3: Comparing the Characteristics of SEDF, and NWECD Workers

	All Workers		Workers earning between $2.50 and $500/hr.		Full-time workers earning between $2.50 and $500/hr.	
	SEDF	NWECD	SEDF	NWECD	SEDF	NWECD
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.465	0.487	0.462	0.484	0.428	0.448
Non-Hispanic white	0.861	0.895	0.862	0.896	0.863	0.895
Black	0.077	0.066	0.076	0.066	0.077	0.067
Hispanic	0.064	0.038	0.063	0.038	0.062	0.037
Ever married	0.761	0.813	0.768	0.819	0.803	0.848
Full-time workers	0.772	0.809	0.784	0.818	1.000	1.000
Occupation						
Manager	0.259	0.257	0.264	0.260	0.281	0.263
Support	0.302	0.242	0.303	0.240	0.293	0.223
Service	0.120	0.096	0.114	0.094	0.088	0.079
Farming	0.021	0.009	0.019	0.009	0.016	0.007
Production	0.118	0.127	0.119	0.129	0.135	0.146
Laborer	0.167	0.239	0.167	0.239	0.173	0.253
Industry						
Agriculture	0.023	0.011	0.022	0.011	0.018	0.009
Mining	0.008	0.007	0.008	0.007	0.009	0.008
Construction	0.062	0.0005	0.062	0.0004	0.067	0.0004
Manufacturing	0.206	0.418	0.209	0.423	0.239	0.470
Transportation	0.077	0.068	0.078	0.069	0.085	0.075
Wholesale	0.047	0.029	0.048	0.029	0.052	0.030
Retail	0.173	0.083	0.169	0.081	0.139	0.065
FIRE	0.070	0.014	0.071	0.014	0.076	0.014
Services	0.334	0.369	0.333	0.366	0.314	0.329
Education						
No high school	0.041	0.037	0.040	0.036	0.037	0.035
Some high school	0.125	0.115	0.120	0.112	0.099	0.102
High school degree	0.316	0.355	0.317	0.356	0.325	0.370
Some college	0.219	0.205	0.219	0.204	0.216	0.197
Associate's degree	0.074	0.090	0.075	0.090	0.078	0.090
Bachelor's degree	0.147	0.125	0.150	0.127	0.159	0.128
Advanced degree	0.077	0.073	0.079	0.074	0.085	0.077
Mean age	38.016	38.887	38.166	39.026	38.440	39.360
	(0.003)	(0.012)	(0.003)	(0.012)	(0.003)	(0.012)
Mean number of	46.359	47.286	46.650	47.521	49.988	50.271
weeks worked	(0.003)	(0.010)	(0.003)	(0.010)	(0.001)	(0.005)
Mean usual hours	39.586	39.624	39.638	39.680	42.261	41.945
worked per week	(0.003)	(0.011)	(0.003)	(0.010)	(0.002)	(0.006)
Mean wage or salary	23,147.38	22,438.53	23,764.47	22,875.89	27,259.79	25,611.50
income	(7.749)	(22.031)	(7.790)	(22.000)	(8.896)	(23.884)
Mean hourly wage	12.617	12.012	12.545	11.983	12.655	11.977
	(0.023)	(0.029)	(0.004)	(0.013)	(0.004)	(0.010)
Number of workers	14,264,082	1,056,635	13,817,006	1,032,462	10,830,247	845,020

Note: The numbers in parentheses are the standard errors of means. The reference period for number of weeks worked, usual hours worked per week, wage or salary income, and hourly wage is the previous year (1989). Hourly wage is estimated as: (wage or salary income/number of weeks worked)/usual hours worked per week.

Table 4: Regressions of Worker Wages for SEDF and NWECD Workers

	Log Hourly Wages			Log Annual Hours		
	SEDF		NWECD	SEDF		NWECD
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.108	-0.108	-0.096	-0.090	-0.090	-0.077
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.003)
Age	0.046	0.046	0.044	0.069	0.069	0.064
	(0.0001)	(0.0001)	(0.0003)	(0.0001)	(0.0001)	(0.0003)
Age^2 x 100	-0.045	-0.045	-0.042	-0.078	-0.078	-0.071
	(0.0001)	(0.0001)	(0.0003)	(0.0001)	(0.0001)	(0.0003)
Ever married	0.193	0.193	0.198	0.127	0.127	0.123
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Black	-0.070	-0.070	-0.067	-0.063	-0.063	-0.059
	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.003)
Hispanic	-0.084	-0.084	-0.072	0.027	0.027	-0.001
	(0.001)	(0.001)	(0.004)	(0.001)	(0.001)	(0.004)
High school degree	0.100	0.100	0.103	0.170	0.170	0.118
	(0.001)	(0.001)	(0.002)	(0.0005)	(0.0005)	(0.002)
Some college	0.168	0.168	0.159	0.154	0.154	0.091
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Associate's degree	0.211	0.211	0.240	0.177	0.177	0.124
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Bachelor's degree	0.369	0.369	0.359	0.189	0.189	0.142
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Advanced degree	0.545	0.545	0.534	0.243	0.243	0.229
	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.003)
Black × female	0.100	0.100	0.063	0.107	0.107	0.090
	(0.001)	(0.001)	(0.004)	(0.001)	(0.001)	(0.004)
Hispanic x female	0.079	0.079	0.062	0.023	0.023	0.020
	(0.001)	(0.001)	(0.004)	(0.001)	(0.001)	(0.005)
Ever married × female	-0.218	-0.217	-0.209	-0.158	-0.158	-0.125
	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.003)
Match	...	-0.026	...	...	-0.002	...
		(0.001)			(0.001)	
R^2	0.350	0.350	0.389	0.208	0.208	0.204
Number obs.	14,264,082	14,264,082	1,056,635	14,264,082	14,264,082	1,056,635

Note: All regressions include controls for region, occupation and industry.