# Examining the Employer-Size Wage Premium in the Manufacturing, Retail Trade, and Service Industries Using Employer–Employee Matched Data

*By* KIMBERLY BAYARD AND KENNETH R. TROSKE*

Economists have long sought to understand why large employers pay higher wages than small employers.[1] Despite decades of study and analysis, researchers have been unsuccessful in fully accounting for the employer-size wage premium. One problem has been insufficient data. Most empirical efforts to examine the size-related wage premium rely on worker-level data that contain little information about a worker's employer, or firm-level data that lack information on the characteristics of a firm's workforce. Although more recent attempts exploit newly developed employer–employee matched data (Kevin T. Reilly, 1995; Troske, 1999), most matched data sets cover only limited segments of the population, thus raising doubt about the generality of the findings.

In this paper, we use the recently created New Worker Establishment Characteristics Database (NWECD) to examine the employer-size wage premium. Unlike other matched data sets, the NWECD contains information on workers and employers in all sectors and regions of the economy. Consequently, we can examine the relationship between employer size and wages separately for workers in different industries. Cross-industry comparisons are important because many explanations of the size-related wage premium emphasize such factors as differences in employers' capital–labor ratio, or workers' computer usage, which vary systematically across industries. We focus here on the three largest industries in our data: manufacturing, retail trade, and services. In addition, we directly examine two recent explanations of the employer-size wage premium: (i) that it reflects productivity differences between workers in large and small establishments (Oi and Idson, 1998), and (ii) that it results from matching more-skilled workers together in large establishments (Michael Kremer and Eric Maskin, 1996; Troske, 1999).

## I. The Data

The NWECD is a cross-sectional data set that links workers' responses to the 1990 decennial Census long form to establishment data drawn from various Economic Censuses. Because this data set is constructed in the same way as the original Worker Establishment Characteristics Database (Troske, 1998) and is documented extensively elsewhere (Bayard et al., 1998), we describe these data only briefly. The NWECD was constructed by matching worker records from the 1990 Sample Detail File (SDF) to establishment records from the 1990 Standard Statistical Establishment List (SSEL). The 1990 SDF consists of all household responses to the 1990 decennial Census long form and contains standard demographic information for all respondents as well as detailed location and industry information for each respondent's place of work. The SSEL is a complete list of all establishments in the United States in a given year and

* Bayard: Department of Economics, University of Maryland, College Park, MD 20742, and U.S. Census Bureau; Troske: Department of Economics, University of Missouri, Columbia, MO 65211, and U.S. Census Bureau. This research was supported by NSF grant SBR95-10876 through the NBER. We thank Peter Mueser for helpful comments. All opinions, findings, and conclusions expressed herein are those of the authors and do not in any way reflect the views of the U.S. Census Bureau or the Center for Economic Studies. The data used in this article were collected under the provisions of Title 13 U.S. Code and are only available at the Center for Economic Studies, U.S. Census Bureau. To obtain access to these data, contact Chief, Center for Economic Studies, U.S. Census Bureau, Room 211/WPII, Washington, DC 20233.

[1] Research on this topic appears to begin with Henry Moore (1911). The employer-size wage premium is extensively examined in Charles Brown and James Medoff (1989) and Brown et al. (1990). For a recent survey of the literature see Walter Y. Oi and Todd L. Idson (1998).

TABLE 1—INDIVIDUAL LOG WAGE REGRESSIONS BY INDUSTRY

| Independent variable | Manufacturing | Retail | Services |
|---|---|---|---|
| Experience | 0.053 | 0.063 | 0.060 |
| | (0.001) | (0.003) | (0.002) |
| (Experience)$^2$ × 10 | −0.019 | −0.025 | −0.025 |
| | (0.001) | (0.002) | (0.001) |
| (Experience)$^3$ × 1,000 | 0.033 | 0.045 | 0.046 |
| | (0.003) | (0.006) | (0.004) |
| (Experience)$^4$ × 100,000 | −0.023 | −0.031 | −0.034 |
| | (0.003) | (0.006) | (0.004) |
| Female | −0.138 | −0.110 | −0.028 |
| | (0.004) | (0.011) | (0.007) |
| Ever married | 0.129 | 0.188 | 0.165 |
| | (0.003) | (0.011) | (0.007) |
| Black | −0.080 | −0.050 | −0.107 |
| | (0.004) | (0.026) | (0.011) |
| Female × black | 0.063 | 0.083 | 0.078 |
| | (0.006) | (0.030) | (0.011) |
| Female × ever married | −0.118 | −0.184 | −0.156 |
| | (0.004) | (0.014) | (0.008) |
| High-school diploma | 0.099 | 0.065 | 0.067 |
| | (0.002) | (0.009) | (0.004) |
| Some college | 0.166 | 0.151 | 0.236 |
| | (0.003) | (0.010) | (0.005) |
| Bachelor's degree | 0.356 | 0.382 | 0.412 |
| | (0.004) | (0.014) | (0.006) |
| Graduate degree | 0.484 | 0.458 | 0.566 |
| | (0.007) | (0.039) | (0.008) |
| Number of children × female | 0.079 | 0.094 | 0.076 |
| | (0.008) | (0.024) | (0.008) |
| Log of establishment employment | 0.047 | 0.050 | 0.047 |
| | (0.003) | (0.008) | (0.004) |
| Log of firm employment | 0.014 | −0.0001 | 0.006 |
| | (0.001) | (0.003) | (0.001) |
| $R^2$: | 0.547 | 0.447 | 0.509 |
| Number of observations: | 348,668 | 22,925 | 141,817 |

*Notes:* Standard errors are in parentheses. Number of children refers to number of children ever born, is asked only of women, and is set to zero for men. Regressions also include controls for four-digit industry, region, occupation, interactions of number of children with experience measures, and whether the plant is located in a MSA. Standard errors are corrected for heteroscedasticity and for the clustered sampling scheme.

is used by the Census Bureau to administer various economic censuses and surveys. The SSEL contains detailed location and industry information for each establishment along with a unique establishment identifier that is common to other Census Bureau economic surveys. Worker records from the SDF are matched to employer records in the SSEL using the common industry and location information for employers. To obtain the information about a worker's employer, we match the NWECD to establishment-level data from the 1987 Censuses of Manufacturing, Retail Trade, and Services.

We impose several restrictions on the data. First, we only keep establishments with 20 or more employees because measures of per-worker earnings taken from the SSEL and SDF are closer for larger establishments than for smaller ones; thus we believe that the data are of higher quality for larger establishments. Second, we only include workers who report usually working more than 30 hours per week and working more than 30 weeks in the previous year. We exclude part-time workers because different patterns of part-time employment across industries may obfuscate comparisons of the size-related wage premium. Third, because we use information from matched workers to estimate characteristics of employers, we limit ourselves to workers in establishments where we match at least 10 percent of the workforce. Finally, to eliminate outliers, we only keep those workers earning between $2.50 and $250.00 per hour and those establishments where the difference between total sales and total payroll is positive.

Evidence presented in Bayard et al. (1998) shows that workers are correctly matched to establishments and that wage regressions using these matched data are unlikely to be biased from sample selection associated with the matching process, even though the data set is nonrepresentative. One of the main sources of this nonrepresentativeness is that the NWECD oversamples workers in manufacturing and undersamples workers in retail trade. This fact provides additional justification for analyzing workers separately by industry.

## II. Empirical Results

Table 1 presents the results from estimating the following model:

$$(1) \qquad \ln W_i = \alpha + \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\gamma} + \varepsilon_i$$

where $W_i$ is the hourly wage of worker $i$, $\mathbf{X}_i$ is a matrix of characteristics for worker $i$ and $\mathbf{Z}_i$ is a matrix of characteristics for worker $i$'s employer. The matrix $\mathbf{X}_i$ contains a quadratic in potential experience and controls for a worker's sex, race, marital status, education, occupation, industry, region, number of children ever born, whether the worker works in an MSA, and interactions of the number of children ever born with experience and sex. The matrix $\mathbf{Z}_i$ includes the log of establishment employment and the log of firm employment.[2]

To compare the employer-size wage premium for workers in industries with much different establishment and firm size distributions, we use a relative, rather than absolute, measure of the impact of size. We compute the establishment-size wage premium as the average wage of workers in establishments with log employment one standard deviation above the mean minus the average wage of workers in establishments with log employment one standard deviation below the mean. The firm-size wage premium is measured analogously. Table 2 presents the cross-worker means and standard deviations for log establishment employment and log firm employment, by industry.

The coefficients in Table 1 show that, even after controlling for demographic characteristics, there is a large and significant establishment-size wage premium in all three industries that is similar to previous estimates of the establishment-size wage premium (Brown and Medoff, 1989). The estimated establishment-size wage premium is 14 percent for workers in manufacturing, 10 percent in retail trade, and 11 percent in services. These estimated premia are quite similar, suggesting that establishment-level factors which differ significantly across industries, such as the capital–labor ratio or computer usage, do not account for much of the establishment-size wage premium. This result is consistent with the finding reported in Troske (1999) that the capital–labor ratio or the use of computers accounts for little of the establishment-size wage premium among manufacturing workers.

TABLE 2—CROSS-WORKER MEAN AND STANDARD DEVIATION OF THE LOG OF ESTABLISHMENT EMPLOYMENT AND THE LOG OF FIRM EMPLOYMENT, BY INDUSTRY

| Variable | Manufacturing | Retail | Services |
|---|---|---|---|
| Mean of log employment in establishment | 6.026 | 4.655 | 6.050 |
| SD of log establishment employment | 1.456 | 0.996 | 1.185 |
| Mean of log firm employment | 8.068 | 9.195 | 6.778 |
| SD of log firm employment | 2.395 | 3.528 | 1.692 |
| Number of observations | 348,668 | 22,925 | 141,817 |

*Note:* We include only observations with positive values for (sales − payroll).

In contrast, there is significant variation in the firm-size wage premium across workers in the three industries. The coefficients on log firm size imply that the firm-size wage premium is 7 percent in manufacturing, 2 percent in services, and nonexistent in retail trade. These cross-industry differences suggest that systematic cross-industry differences in firm characteristics can account for the firm-size wage premium.

We now turn to examining two possible explanations for the size-related wage premium. The first hypothesis, presented by Oi and Idson (1998), states that the premium reflects the greater productivity of workers in large establishments or firms due to differences in the organization of production, the quality of capital, the level of effort required by employers, or the amount of employer-specific training. To examine this hypothesis we include in $\mathbf{Z}_i$ a measure of labor productivity in a worker's establishment. We define establishment-level productivity as [(total sales − total payroll)/total hours worked].[3] The second hypothesis

[2] See Oi and Idson (1998) for a discussion of why employment is the appropriate measure of size.

[3] Total hours worked is not available for nonproduction workers in the Census of Manufacturing, or for any workers in the Censuses of Retail Trade and Services. We

we examine is that employers prefer to match workers of similar skill together within establishments (Kremer and Maskin, 1996). If there are fixed costs associated with hiring more-skilled workers, then more-skilled workers will be matched together in large establishments (Troske, 1999). To investigate this hypothesis we include in $\mathbf{Z}_i$ measures of the skill of the workforce in a worker's establishment. We employ three measures of skill: mean years of potential experience of workers in the establishment, the percentage of workers in an establishment with some postsecondary education but not a college degree, and the percentage of workers in an establishment with at least a college degree. In both of these models, the characteristics of employers are capturing unobserved worker skill.

Table 3 presents the regression results from examination of the two explanations for the size-related wage premium. The results show that workers in all three industries who work in more productive establishments, and who work in establishments with a more skilled workforce, earn higher wages. However, the coefficients on log establishment employment reported in Table 3 are approximately the same as those in Table 1, indicating that the productivity and skill measures have almost no effect on the estimated establishment-size wage premium. In contrast, including the productivity and skill measures in the regression does reduce the estimated firm-size wage premium from 7 percent to 3 percent for manufacturing workers, and from 2 percent to 1 percent for service workers (according to calculations based on Tables 1 and 3). Results based on adding the labor-productivity and workforce-skill measures separately (available from the authors upon request) show that the labor-productivity measure accounts for almost all of this decline. Thus, it appears that a significant portion of the firm-size wage premium is the result of employees working in more productive establishments.

TABLE 3—INDIVIDUAL LOG-WAGE REGRESSIONS INCLUDING PRODUCTIVITY AND WORKFORCE SKILL

| Independent variable | Manufacturing | Retail | Services |
|---|---|---|---|
| Log of establishment employment | 0.044 (0.003) | 0.047 (0.007) | 0.047 (0.003) |
| Log of firm employment | 0.007 (0.001) | 0.0005 (0.003) | 0.0017 (0.0014) |
| Log[(sales − payroll)/(total hours)] | 0.056 (0.003) | 0.055 (0.011) | 0.050 (0.006) |
| Mean experience in establishment | 0.006 (0.0005) | 0.003 (0.001) | 0.003 (0.001) |
| Percentage of workforce with some college | 0.210 (0.012) | 0.099 (0.024) | 0.059 (0.020) |
| Percentage of workforce with college degree | 0.155 (0.020) | 0.154 (0.042) | 0.165 (0.029) |
| $R^2$: | 0.557 | 0.451 | 0.511 |
| Number of observations: | 348,668 | 22,925 | 141,817 |

*Notes:* Standard errors are in parentheses. All regressions also include the same controls as in Table 1. All standard errors have been corrected for heteroscedasticity and for the clustered sampling scheme.

## III. Conclusion

The estimated establishment-size wage premium is strikingly similar across the three industries, while the estimated firm-size wage premium is quite different. In this paper, we examine whether greater productivity or higher concentration of skilled workers in large establishments and firms explains the size-related wage premium. We find that neither productivity nor segregation by skill accounts for the premium across individual establishments. However, we do find that the greater productivity of workers in larger establishments does account for over half of the firm-size wage premium in both manufacturing and services.

Our inability to account for the establishment-size wage premium suggests that there may be complex differences in the organization of production between large and small employers

compute total hours worked in an establishment by first constructing the average of total hours worked for all workers (part-time and full-time) matched to a particular establishment and then multiplying this average by the total number of workers in the establishment. Additionally, we run the regressions with productivity controls using an alternative measure of labor productivity [(sales − payroll)/(total employment)] and obtain results nearly identical to those in Table 3.

that are not captured by our simple measures of productivity and workforce skill. Oi and Idson (1998) hypothesize that one difference between large and small employers is that large employers simply have more output over which to amortize large sunk-cost investment. This gives large employers a greater ability to attract and retain high-ability workers, to organize workers into teams, and to use higher-quality capital equipment. Clearly, the employer-size wage premium reflects more fundamental differences between large and small firms than how they compensate workers, and understanding it requires a more complete grasp of how differences in firm size affect the way firms organize to produce output.

## REFERENCES

**Bayard, Kimberly; Hellerstein, Judith; Neumark, David and Troske, Kenneth.** ''Why Are Racial and Ethnic Wage Gaps Larger for Men Than for Women? Exploring the Role of Segregation Using the New Worker–Establishment Characteristics Database.'' Mimeo, University of Missouri-Columbia, 1998.

**Brown, Charles; Hamilton, James and Medoff, James.** *Employers large and small*. Cambridge, MA: Harvard University Press, 1990.

**Brown, Charles and Medoff, James.** ''The Employer Size–Wage Effect.'' *Journal of Political Economy*, November 1989, *97*(5), pp. 1027–59.

**Kremer, Michael and Maskin, Eric.** ''Segregation by Skill and the Rise in Inequality.'' National Bureau of Economic Research (Cambridge, MA) Working Paper No. 5718, August 1996.

**Moore, Henry L.** *Laws of wages: An essay in statistical economics*. New York: Augustus M. Kelley, 1911.

**Oi, Walter Y. and Idson, Todd L.** ''Firm Size and Wages.'' Mimeo, University of Rochester, 1998.

**Reilly, Kevin T.** ''Human Capital and Information.'' *Journal of Human Resources*, Winter 1995, *30*(1), pp. 1–18.

**Troske, Kenneth R.** ''The Worker Establishment Characteristics Database,'' in John Haltiwanger, Marilyn Manser, and Robert Topel, eds., *Labor statistics measurement issues*. Chicago, IL: University of Chicago Press, 1998.

_____. ''Evidence on the Employer-Size Wage Premium from Worker–Establishment Matched Data.'' *Review of Economics and Statistics*, February 1999, *81*(1), pp. 1–12.