Using the National Compensation Survey to Predict Wage Rates

The National Compensation Survey (NCS) combines three former Bureau of Labor Statistics' establishment survey programs: The Employment Cost Index, the Employee Benefits Survey, and the Occupational Compensation Survey. As part of the NCS, the Bureau collects data on job characteristics, dubbed "generic leveling factors." This study demonstrates the usefulness of these data in predicting wage rates.

BROOKS PIERCE

Brooks Pierce is an economist in the Compensation Research and Program Development Group, Bureau of Labor Statistics. The results presented here rely heavily on material produced by Maury Gittleman, an economist in the same office.
Telephone: (202) 691-7525
E-mail: Pierce_B@bls.gov

The National Compensation Survey (NCS) is designed to be a nationally representative random sample of jobs in the non-agricultural, non-Federal economy. Field economists visit sampled establishments and obtain information on the establishment and a sample of jobs in the establishment. At present the data collected are mainly attributes of the job and establishment, and earnings and work schedules. Establishment data include employment size and location, the establishment's industry according to the Standard Industrial Classification (SIC) system, and whether the establishment is privately owned or operated by a State or local government. Occupation-specific data include Bureau of the Census occupational classifications, whether the occupation is covered by a union contract, whether the job is part time or full time, and generic leveling factors data. Earnings data, including incentive pay data, are collected for the individuals in a sampled occupation; per hour wage rates are calculated in conjunction with work schedule information for the occupation.

The generic leveling factors are designed to measure job duties. There are at present 10 factors, each with various levels: Knowledge (9 levels), supervisory controls (5), guidelines (5), complexity (6), scope and effect (6), personal contacts (4), purpose of contacts (4), physical demands (3), work environment (3), and supervisory duties (5). The factor names largely reflect their content. For example, part of the description of the first level of knowledge is "...simple, routine, or repetitive tasks or operations which typically include following step-by-step instructions and requires little or no previous training or experience." In comparison, the fourth level of knowledge is described in part as "Knowledge of an extensive body of rules, procedures, operations, products or services requiring extended training and experience to perform a wide variety of interrelated or nonstandard procedural assignments and resolve a wide range of problems."

These data elements are "generic" in the sense that they do not rely on identifying the occupation in question. This facilitates the collection of these data for random samples of jobs that cover the broad range of occupations in the economy. It also gives some basis for comparing or classifying occupations that are distinct but that may have similar duties and responsibilities. Although designed to describe the job and not the workers in the job, these factors are in practice likely to be related to job incumbent abilities, for the simple reason that employers recruit workers in such a way as to match employee abilities with job duties.

The sample used in this article corresponds to that used for the 1997 national NCS summary publication. The sample includes 145,054 jobs from 17,246 different establishments.

TABLE 1. **Employment distribution by generic leveling factor across levels, 1997 National Compensation Survey**
(Percent)

Generic leveling factor	Total	Level								
		1	2	3	4	5	6	7	8	9
Knowledge	100	13.1	28.0	18.7	11.8	6.6	13.9	6.3	1.5	0.1
Supervision received	100	25.6	40.1	27.0	6.3	1.0	-	-	-	-
Guidelines	100	37.7	34.1	22.6	4.8	.7	-	-	-	-
Complexity	100	22.6	35.1	33.1	6.3	2.8	.2	-	-	-
Scope and effect	100	34.6	33.3	26.5	4.3	1.2	.1	-	-	-
Personal contacts	100	48.1	40.0	11.6	.3	-	-	-	-	-
Purpose of contacts	100	64.6	26.7	8.1	.7	-	-	-	-	-
Physical demands	100	40.7	56.6	2.8	-	-	-	-	-	-
Work environment	100	49.2	48.9	1.9	-	-	-	-	-	-
Supervisory duties	100	81.6	7.9	8.7	1.7	.2	-	-	-	-

NOTE: Dash indicates data not applicable. Due to rounding, sums of percentages may not equal 100.

Employment distribution by generic leveling factors

Table 1 shows employment distribution by generic leveling factors. These distributions are silent on whether jobs are properly classified and on whether each level is clearly distinct from its neighbors. Nevertheless, very sparse data cells might suggest that too-fine distinctions are being made, whereas cells with excessive amounts of data would suggest that the scale as constituted is not very useful in distinguishing one job from another.

For most of the factors, the scales do not indicate such data problems. The knowledge scale has the most levels, by design, as wages are known to be closely related to factors this scale captures, and because it is probably easier to make meaningful distinctions along this dimension. For this factor and the next four, jobs are widely dispersed among the categories. The remaining five factors tend to have more clumping of the data. This largely reflects the fact that many jobs simply do not have duties associated with the factor in question. For example, most jobs are not supervisory in nature, so the masses of the data are in the lower levels of supervisory duties. Many jobs are sedentary and have low physical demands (the physical demands scale is from less to more exertion), and most work environments involve only everyday risks (the work environment scale is from low risk to high risk). It is possible that collecting some of these data in less detail would reduce respondent burden without unduly reducing the factors' predictive power for wages. Whether or not the scale distinctions for, say, supervisory duties, can help predict wages independently of the other factors is an issue this article addresses.

Correlation among factors

One aspect of these data that is not obvious from table 1 is the question of how independent one factor's information is from the others'. That is, does knowing a job's complexity level tell you much more about the job than if you had only known the job's knowledge level? This is an important question because its answer gives some idea about how redundant the information incorporated in, for example, "complexity" is to that incorporated in "knowledge." Indeed, if position in the complexity scale can be completely and accurately determined from position in the knowledge scale, there is no reason to construct the complexity scale at all.

Table 2 gives rank order correlation coefficients between the various factors to address this general question. The most striking result here is the large positive correlation among the first seven factors, which suggests some duplication of data. Therefore, there may be difficulty in disentangling the effects of knowledge on wages from the effects of, say, complexity on wages. However, despite the high correlation, each factor may still contribute to explaining wages or otherwise classifying jobs because each factor may embody information not contained in the other factors. The supervisory duties variable is somewhat less correlated with knowledge. The physical demands and work environment variables are positively correlated with each other, and are somewhat negatively correlated with knowledge and the other factors.

Many of the factors seem related to knowledge. As an example, table 3 compares distributions of complexity levels conditional on different levels of knowledge. This gives some indication of the interrelated nature of the factors. Complexity is chosen only as an example; the qualitative point is similar for other factors. Over 95 percent of sampled jobs with knowledge level 1 also have the lowest level of complexity. Clearly, higher levels of knowledge are associated with higher levels of complexity. For instance, at knowledge level 2 about 35 percent of

TABLE 2. **Rank order correlation coefficents of generic leveling factors, 1997 National Compensation Survey**

Generic leveling factor	Knowl-edge	Super-vision received	Guide-lines	Com-plexity	Scope and effect	Personal contacts	Purpose of contacts	Physical demands	Work environ-ment	Super-visory duties
Knowledge	1.00	-	-	-	-	-	-	-	-	-
Supervision received	.83	1.00	-	-	-	-	-	-	-	-
Guidelines	.82	.85	1.00	-	-	-	-	-	-	-
Complexity	.84	.85	.85	1.00	-	-	-	-	-	-
Scope and effect	.82	.84	.87	.85	1.00	-	-	-	-	-
Personal contacts	.72	.64	.60	.60	.60	1.00	-	-	-	-
Purpose of contacts	.74	.67	.65	.64	.64	.74	1.00	-	-	-
Physical demands	-.45	-.32	-.30	-.31	-.25	-.51	-.41	1.00	-	-
Work environment	-.33	-.21	-.18	-.18	-.14	-.47	-.35	.78	1.00	-
Supervisory duties	.47	.47	.45	.44	.44	.39	.46	-.16	-.12	1.00

NOTE: Dash indicates redundant data.

the jobs have a complexity level of 1, and most are classified in the next higher level. The dispersion of complexity within knowledge levels indicates how useful that factor could be in predicting wages independently of knowledge. To take the most extreme example, one wouldn't expect complexity to explain much of the observed wage variance in a sample of knowledge level 1 jobs, simply because there is little complexity variation in that sample. On the other hand, complexity could potentially explain a great deal of the observed wage variance in a sample of knowledge level-5 jobs.

Generic leveling factors as wage determinants

Wage rates vary substantially by industry, location, occupation, union status, full- and part-time employment, and establishment employment size.[1] These wage predictors explain much but not all of the observed wage variation. This section compares the predictive power of the generic leveling data to that of other wage predictors.

Table 4 gives R-squared values for a battery of log wage regressions.[2] Statistics are presented first for more traditional variables and the less traditional generic leveling factors data taken together and then for each set separately. The first number in table 4, .847, indicates that almost 85 percent of the wage variation in these data can be explained by the generic leveling factors in combination with the more traditional covariates.

Controls for the traditional job and establishment attributes, without the generic leveling factors, explain 72.6 percent of the wage variation. Occupation is the most important of these variables; controls for occupational groups alone can explain about 59 percent of the wage variation in these data.[3] Other factors that alone tend to explain a substantial amount of wage variation are industry groups (as measured by 2-digit SIC indicators), establishment size (as measured by employment), and ownership (whether the establishment is private or operated by a State or local government). On the other hand, union status has little explanatory power in this R-squared sense. Since most jobs are nonunion, and because there are so many factors affecting wage rates, unionization by itself explains a small fraction of overall wage variation.

Table 4 also gives R-squared values for a series of regressions where the explanatory variables are indicator variables for the generic leveling factors. The R-squared for knowledge (.702) indicates that this factor alone explains almost as much wage variance as do all of the traditional covariates (.726). On the other hand, some of the other generic leveling factors do not explain very much of the overall wage variance. For example, supervisory duties alone would explain about 18 percent of the wage variance. The same explanation applies here as with the union status variable. Even if there are large differences in wages for different levels of supervisory duties, that variable explains a relatively small portion of wage variance because there is only a small amount of variation in the explanatory variable and a large amount of variation in wages.

The generic leveling factors explain about 75 percent of the wage variance, somewhat more than the traditional covariates at about 73 percent. If one had to choose between the set of traditional covariates and the set of generic leveling variables, with the sole purpose of maximizing predicted wage variance, one would choose the generic leveling data. Given that combining the generic leveling data and the traditional covariates in the same regression raises the R-squared to .847, it appears that each set of variables has something to offer that the other does not, and that they at least partially measure different things.

Looking beyond the R-squared statistics, the regressions indicate the wage differentials associated with each explanatory variable. For example, they show the measured union wage premium when other factors are controlled for. Or, the regressions show the wage premium associated with jobs having knowledge level 2 rather than knowledge level 1, controlling for other factors. Table 5 gives some in-

TABLE 3. **Employment distribution by complexity levels across knowledge levels, 1997 National Compensation Survey**

(Percent)

Complexity level	Knowledge level 1	2	3	4	5	6	7	8	9
1	95.4	35.2	0.9	0.1	0.3	0	0	0	0
2	4.6	61.9	60.9	21.7	33.8	7.2	.1	0	0
3	0	2.9	38.1	74.3	50.4	85.8	18.5	.2	0
4	0	0	.1	3.9	15.5	6.9	59.0	6.5	1.6
5	0	0	0	0	0	.1	21.9	88.3	19.3
6	0	0	0	0	0	0	.4	5.0	79.2
Knowledge marginal distribution	13.1	28.0	18.7	11.9	6.6	13.9	6.3	1.5	.1

NOTE: The first six rows give the distribution of the complexity factor conditional on having the given value for knowledge. For example, 95.4 percent of knowledge level = 1 jobs have complexity level = 1, and 4.6 percent of knowledge level = 1 jobs have complexity level = 2. Hence the numbers in the first six rows sum to 100 percent for each column. The final row gives the percent of jobs with the given level of knowledge. For example, 13.1 percent of the jobs have knowledge level = 1. Therefore the final row sums to 100 percent across columns. Due to rounding, sums of percentages may not equal 100.

TABLE 4. **Variance results by job and establishment attributes and generic leveling factors, 1997 National Compensation Survey**

Job and establishment attributes plus generic leveling factors	R-squared[1]
All job and establishment attributes plus generic leveling factors	.847
All job and establishment attributes	.726
Occupational group	.590
Full-time status	.132
Industry group	.303
Establishment employment size	.089
Ownership	.050
Survey locality	.069
Union status	.041
Incentive pay	.000
All generic leveling factors	.749
Knowledge	.702
Supervision received	.626
Guidelines	.623
Complexity	.625
Scope and effect	.605
Personal contacts	.386
Purpose of contacts	.430
Physical demands	.170
Work environment	.076
Supervisory duties	.179

[1] Proportion of wage variance explained by the indicator variables.

NOTE: Statistics are R-squared numbers from regressions of log wages on the indicated characteristics. With the exception of employment size, all job and establishment attributes are represented by indicator variables. Establishment employment size is continuous and is entered in log form. The occupational classification has 42 groups; the industry classification has 75 groups. Generic leveling data are entered as indicator variables for each level of the separate factors.

dication of the premiums attached to a subset of these variables. Two regressions are presented. The first includes the usual job and establishment attributes. The second includes all generic leveling factors as well as the usual job and establishment attributes. In subsequent tables, the former regression is referred to as "including job and establishment attributes" and the latter is referred to as "including job and establishment attributes plus generic leveling factors."

As expected, wage rates are positively related to establishment size, union status, full-time status, and incentive pay, even controlling for other factors. Because the dependent variable in these regressions is the natural logarithm of the hourly wage rate, coefficients may be interpreted as approximate percentage differentials. The union status, full-time status, and incentive pay variables are all indicator variables (meaning they take the value 1 when the category applies, and 0 when it does not), so that the interpretation of the coefficients as approximate percentage differentials is straightforward. For example, the union status coefficient of .168 indicates that wages in union covered jobs are about 17 percent higher than those in nonunion jobs, after controlling for the other covariates listed.[4] Because the establishment size variable is continuous, it may help to report the predicted wage difference associated with some fixed difference in establishment employment size. A one standard deviation change in the natural logarithm of establishment employment is 1.67, which implies a predicted wage change of 1.67 times .030, or about 5 percent.[5] The wage premiums listed in both columns are fairly large, and all are estimated precisely (as evidenced by relatively small standard errors). Note that the coefficients from the regression including job and establishment attributes are generally smaller than the corresponding ones from the regression including job and establishment attributes plus generic leveling factors, meaning that controlling for differences in the generic leveling factors tends to reduce the estimated wage premiums of other variables.[6]

Table 5 does not present all of the wage coefficient estimates. However, because the effects of the generic leveling data are of some independent interest, those results are given in table 6. The regression including job and establishment attributes plus generic leveling factors has 40 variables that describe the generic leveling information—each generic leveling factor has a set of indicator variables associated with it. Estimated coefficients and standard errors for these 40 variables are presented on a factor by factor basis in table 6. In all cases, the variables describe the effect on wages of having a higher level of some factor, as opposed to having the lowest possible level of that factor. For example, the first number in table 6 is 0.089, for the second level of knowledge. This indicates that jobs with knowledge level 2 are estimated to have wages about 9 percent higher than otherwise comparable jobs with knowledge level 1.[7] The standard error of 0.009 indicates that the 9 percent point estimate is rather precisely estimated,

TABLE 5. **Wage regression coefficient estimates and standard errors for selected job and establishment attributes, 1997 National Compensation Survey**

Job and establishment attribute	Regression includes— Job and establishment attributes		Job and establishment attributes plus generic leveling factors	
	Coefficient	Standard error	Coefficient	Standard error
Log of establishment employment size	0.030	0.002	0.021	0.002
Union status	.168	.008	.160	.006
Full-time status	.234	.006	.120	.005
Incentive pay	.159	.015	.107	.011

NOTE: The dependent variable is the average log wage of incumbent workers in the sampled job. Both regressions include indicator variables for ownership, detailed industry classification (2-digit SIC), detailed occupational group, and survey locality. The second regression also includes indicator variables for levels of each generic leveling factor. Standard errors are robust, and include corrections for clustering by establishment. Within 17,246 establishments, 145,054 observations were made.

and is statistically different from zero. The coefficient estimate of 0.171 for knowledge level 3 means that wages among those jobs are approximately 17 percent higher than are wages among otherwise comparable jobs with knowledge level 1. The approximate premium for moving from knowledge level 2 to knowledge level 3 is about 8 percent (0.171-0.089 = 0.082).[8]

In fact, the estimated premiums from increasing knowledge one level is routinely on the order of 8 to 15 percent, except at the very top end of the knowledge scale. These premiums are comparable to or slightly lower than those for union status, full-time status, and incentive pay. For example, the full-time status wage premium of 0.120 from table 5 is about equal to the wage differential associated with having knowledge level 7 rather than knowledge level 6.

The other generic leveling factors tend to exhibit lower wage premiums, and some factors tend not to be very useful in predicting wages once other factors are controlled for. The wage premium for an additional level of supervisory control is about 8 percent throughout its range. The estimated premium associated with a higher level of guidelines is about 5 percent at lower levels and somewhat more at higher levels. The analogous estimates for complexity are 3 to 5 percent at lower levels and closer to 10 percent at the highest level, although those effects are less precisely estimated than some of the other coefficients. In this regression, the scope and effect, personal contacts, purpose of contacts, work environment, and physical demands factors tend to have small and sometimes imprecisely estimated coefficients. The wage return to higher supervisory duties is small at lower levels but fairly substantial at higher ones. In interpreting these results, note that the very top levels of these factors are sparsely populated. The larger standard errors for the top-level coefficients reflect that sparseness. Nevertheless, it appears that several of the generic leveling factors are useful in predicting wages, at least as compared to the more traditional covariates. For example, the wage premium associated with two extra levels of supervisory control is about the same as that associated with union status.

Between- and within-occupational variation

Generic leveling data are meant to capture differences across jobs in job duties and tasks. Do occupational identifiers also capture those differences across jobs? A comparison of the data in table 5 shows that job and establishment attributes combined with generic leveling factors contain data not captured by job and establishment attributes alone. Just how do the generic leveling factors relate to occupation?

Table 5 gave results from two different regressions. Both regressions control for job and establishment attributes, which include indicator variables for occupational group. The difference between the two regressions is that one controls for the generic leveling factors while the other does not. If the generic leveling factors and occupational indicators both capture differences across jobs in job duties, then the coefficients on the occupational indicator—which are interpretable as occupational wage premiums—are likely to be quite different in the two regressions. For example, engineering jobs may pay quite a bit more than other jobs, but they may not pay more than other jobs with comparable levels of knowledge and other generic leveling factors.

Table 7 gives the estimated occupational log wage premiums corresponding to each of these two regressions. The statistics in the column labeled "Premium" describe wage differences between occupations. The first number in table 7, .620, is the estimated occupational premium for public administration officials from the regression including job and establishment attributes. This statistic reflects a very large wage rate for public administration officials relative to the average occupation, even after controlling for the traditional covariates in table 5.[9] The premium of -.080 for the same occupation from the regression including job and establishment attributes plus the generic leveling factors indicates that public administration officials actually are estimated to earn about 8 percent less than the average occupation, once the generic leveling factors are also controlled for. This implies that much of the difference in wage rates between the average occupation and public administration jobs is attributable to differences in the generic leveling factors. In fact, most of the estimated wage differentials shrink dramatically once controls are instituted for the generic leveling factors. This is reassuring from the standpoint of data collection, because it suggests that the generic leveling

TABLE 6. **Wage regression coefficient estimates and standard errors for generic leveling factor levels, 1997 National Compensation Survey**

Coefficient and standard error by generic leveling factor level	Generic leveling factor									
	Knowledge	Supervision received	Guidelines	Complexity	Scope and effect	Personal contacts	Purpose of contacts	Physical demands	Work Environment	Supervisory duties
1										
Coefficient	-	-	-	-	-	-	-	-	-	-
Standard error	-	-	-	-	-	-	-	-	-	-
2										
Coefficient	0.089	0.081	0.060	0.029	0.029	-0.012	0.027	-0.023	0.036	0.017
Standard error	.009	.009	.007	.006	.007	.005	.005	.007	.007	.005
3										
Coefficient	.171	.156	.104	.078	.053	.018	.038	-.017	.062	.059
Standard error	.012	.011	.010	.009	.009	.009	.009	.013	.017	.009
4										
Coefficient	.302	.234	.173	.096	.074	.071	.027	-	-	.116
Standard error	.012	.015	.014	.013	.013	.031	.027	-	-	.013
5										
Coefficient	.451	.318	.265	.150	.162	-	-	-	-	.320
Standard error	.016	.028	.030	.018	.031	-	-	-	-	.033
6										
Coefficient	.590	-	-	.265	.078	-	-	-	-	-
Standard error	.018	-	-	.055	.049	-	-	-	-	-
7										
Coefficient	.713	-	-	-	-	-	-	-	-	-
Standard error	.021	-	-	-	-	-	-	-	-	-
8										
Coefficient	.818	-	-	-	-	-	-	-	-	-
Standard error	.033	-	-	-	-	-	-	-	-	-
9										
Coefficient	.806	-	-	-	-	-	-	-	-	-
Standard error	.059	-	-	-	-	-	-	-	-	-

NOTE: This table presents estimated coefficients for generic leveling indicator variables from the log wage regression including job and establishment attributes plus generic leveling factors. Standard errors are robust, and include corrections for clustering by establishment. Dash indicates data not applicable.

data, in large part, measure one of the things they were designed to: Differences across occupations in job duties.

One conclusion to be drawn from table 7 is that occupational detail and generic leveling data to some extent represent substitute information. Is the generic leveling information also valuable for within-occupation wage determination? We know that knowledge and the other factors help predict wage differences between, say, engineers and sales representatives, but do they also help us predict wage rates among engineers?

The variance columns in table 7 address within-occupational variances. These statistics are constructed as follows. The wage regressions from table 5 imply predicted log wage rates for each observation in the data. The actual and predicted wage rates will of course differ since the regressions do not predict perfectly. The difference between the actual and predicted wage rate is usually termed a "residual" or unexplained component. The statistics in the table 7 column labeled "Variance" are the variances of that residual component among the observations in each occupational category. The regressions differ only in that one includes the generic leveling factors as explanatory variables and the other does not. If the generic leveling factors help predict wage differences within occupational group, then the residual components based on the regression including job and establishment attributes plus generic leveling factors will have smaller variances. Among public administration officials, for example, the residual wage variance is .149 based on the regression that does not include generic leveling factors and .077 based on the regression that does. That is, even among like jobs the generic leveling factors help explain a substantial amount of wage variation.

The regression results listed thus far employ an occupational classification scheme with about 40 groups. It is also possible to investigate within-occupational wage variation using a finer occupational classification. To carry out that investigation a very detailed occupational classification system is used, the 1990 census system, which has approximately 470 occupations. Table 8 shows the knowledge distributions within the 10 most populous of these occupations. For example, about 2.6 percent of the data are for registered nurses, and 75 percent of registered nurses are coded as having knowledge level 6. The knowledge variance within occupation is smaller than the knowledge variance in the sample as a whole. Nonetheless, there are some within-occupation differences. Most of the occupations

TABLE 7. **Occupational wage premiums and within-occupation residual wage variances, 1997 National Compensation Survey**

Occupation	Regressions include— Job and establishment attributes		Regressions include— Job and establishment attributes plus generic leveling factors	
	Premium[1]	Variance[2]	Premium[1]	Variance[2]
Executive, administrative, managerial				
Public administration officials	0.620	0.149	-0.080	0.077
Other executives and managers	.774	.162	.003	.056
Management-related occupations	.358	.102	-.029	.042
Professional specialty occupations				
Engineers, architects and surveyors	.606	.077	.014	.033
Mathematical and computer scientists	.644	.105	.062	.041
Natural scientists	.438	.123	-.114	.057
Health diagnosing occupations	1.00	.557	.094	.293
Health assessment and treating	.539	.066	.062	.045
Postsecondary teachers	.871	.146	.109	.097
Other teachers	.550	.095	.158	.066
Lawyers and judges	.842	.169	.027	.070
Other professional specialty	.333	.130	-.066	.063
Technicians				
Health technologists and technicians	.125	.076	.026	.039
Engineering and science technicians	.163	.076	.028	.032
Other technicians	.343	.240	.090	.111
Sales occupations				
Sales supervisors and proprietors	.277	.182	-.007	.077
Sales, finance and business	.215	.258	.073	.113
Sales, commodities excluding retail	.386	.182	.104	.104
Sales, retail and personal services	-.247	.076	-.032	.049
Other sales-related occupations	-.160	.092	-.029	.040
Administrative support				
Supervisors, administrative support	.215	.060	-.068	.040
Computer equipment operators	-.061	.065	.032	.034
Secretaries, typists	-.061	.057	.053	.033
Financial records processing	-.112	.056	.031	.031
Mail and message distributing	-.373	.061	.010	.033
Other administrative support	-.213	.061	-.001	.035
Service occupations				
Protective service	-.049	.116	-.030	.047
Food service	-.439	.139	-.159	.115
Health service	-.322	.048	-.076	.032
Cleaning and building service	-.358	.057	-.028	.039
Personal service	-.303	.128	-.076	.061
Precision production, craft, repair				
Mechanics and repairers	.090	.069	.018	.037
Construction trades	.109	.076	.035	.039
Other precision production	-.041	.111	-.003	.037
Operators, fabricators, laborers				
Machine operators and tenders	-.247	.067	-.002	.039
Fabricators, assemblers, inspectors	-.297	.075	-.002	.054
Motor vehicle operators	-.120	.071	.091	.050
Other transport and material moving	-.123	.075	.065	.045
Construction laborers	-.319	.072	.064	.040
Freight, stock, and material handlers	-.327	.057	.006	.044
Other handlers and laborers	-.376	.073	-.021	.045
Farm-related occupations	-.234	.069	.041	.030

[1] Occupational wage premium. Log wage differentials between the given occupation and the average occupation.

[2] Within-occupation residual wage variance. Variance of regression residuals within the stated occupational category.

TABLE 8. **Employment distribution by knowledge level across selected occupations, 1997 National Compensation Survey**

Knowledge level	Selected occupation									
	Registered nurses	Janitors and cleaners	Cashiers	Elementary school teachers	Nursing aids, orderlies and attendants	Secretaries	Managers and administrators (n.e.c.)	General office clerks	Assemblers	Stock handlers and baggers
1	0	55.0	37.2	0	5.1	0.1	0	7.0	24.0	56.8
2	0	36.1	54.3	.1	72.7	13.5	0	41.9	61.1	38.1
3	.3	7.9	7.9	0	20.8	57.5	.1	43.1	13.5	4.8
4	1.4	1.0	.4	.1	1.2	28.5	.9	7.9	1.3	.3
5	16.4	0	.2	18.4	.2	.4	6.2	.1	.1	0
6	75.0	0	0	79.5	.1	0	31.6	0	0	0
7	6.8	0	0	1.9	0	0	42.9	0	0	0
8	.1	0	0	0	0	0	17.2	0	0	0
9	0	0	0	0	0	0	1.1	0	0	0
Percent of workers in occupation[1]	2.59	2.51	2.50	2.31	2.31	2.21	2.01	2.00	1.80	1.72

[1] Percent of the 1997 National Compensation Survey sample weight in the given occupation.

NOTE: n.e.c. means "not elsewhere classified."

listed in table 8 have a clear modal knowledge level (that is, one knowledge level stands out). Typically, neighboring knowledge levels are substantially populated as well. Although based on a few sample occupations, table 8 suggests non-trivial within-occupational differences in the generic leveling data.

Table 9 shows that the within-occupation differences in the generic leveling data can explain some of the within-occupation wage differences. It shows the results of comparing the variance of predicted wages within a census occupational category to the variance in actual wages within the same category. Predicted wages are based on a regression of log wages on census occupational and generic leveling indicators.[10] Because the generic leveling factors do not explain wages perfectly, the variance in predicted wage is smaller than the variance in actual wages. Therefore, the ratio of the predicted wage variance to the actual wage variance within the occupation forms an index of how well the generic leveling data predict wages within that occupation. This ratio is averaged for all occupations taken together, as well as for broad occupational categories. For all 468 census occupations in the sample, the generic leveling data explain, on average, 42 percent of the within-occupation wage variance.

A final question relates to the nature of the occupations for which the generic leveling data predict wages well. One would not necessarily think that these data would be as helpful in predicting wages within, for example, a group of janitors, as within a group of accountants. The nature of the generic leveling data lead one to suspect that they are more able to distinguish among similar white-collar workers than to distinguish among similar blue-collar workers.[11] For professional occupations, on average, about half of the wage variation within a very narrowly defined occupation can be accounted for by variation in the generic leveling data. The analogous fractions for technical occupations and executive occupations are quite a bit higher than even that. On the other hand, the generic leveling data explain a small fraction of the within-occupation wage variation for blue-collar and service occupations. This indicates that the generic leveling data is more useful in distinguishing among white- than blue-collar or service sector jobs, at least when defining jobs using the census occupational classification system.

Summary

- Jobs with higher levels of generic leveling factors tend to have substantially higher wage rates. This is especially true for cognitive and managerial-related factors.

- Along with covariates traditionally used in labor economics research (such as industry, union or full-time status, etc.), generic leveling data explain a high proportion of the observed wage variation.

- When included as additional controls in wage equations, generic leveling data tend to be lower than the estimated wage differentials associated with some of the more traditional establishment survey wage determinants, such as establishment size and industry differentials.

- Additionally, generic leveling data explain wage differences across occupations very well, and explain more of the wage differences within white-collar occupations than within blue-collar occupations. ■

TABLE 9. **Mean fraction of within-occupation wage variance explained by generic leveling factors by occupational categories, 1997 National Compensation Survey**

Occupational category	Number of occupations	Mean fraction
All occupation average	468	.42
White collar		
Professional	100	.48
Technical and related	22	.65
Executive, administrative, and managerial	26	.58
Sales	22	.39
Administrative support, including clerical	54	.42
Blue collar		
Precision production, craft, and repair	98	.46
Machine operators, assemblers, and inspectors	56	.39
Transportation and material moving	24	.30
Handlers, equipment cleaners, helpers, and laborers	27	.27
Service		
Service, except private household	39	.35

NOTE: Data in this table are derived from a log wage regression on 3-digit census and generic leveling indicator variables. For each 3-digit census occupation, the fraction of the within-occupation wage variance explained by the generic leveling factors is defined as the variance of the regression predicted values divided by the variance of the actual log wage rates. The column labeled "mean fraction" presents the weighted average of this fraction across the occupations in the stated occupational group, with weights determined by NCS sampling weights.

[1] Classic references include Charles Brown and James Medoff, "The Employer Size-Wage Effect," 97(5), *Journal of Political Economy*, October 1989, pp.1027-1059; Alan B. Krueger and Lawrence H. Summers, "Efficiency Wages and the Wage Structure," *Econometrica*, 56(2), March 1988, pp. 259-293; and H. Gregg Lewis, *Union Relative Wage Effects: A Survey*, University of Chicago Press, Chicago, 1986.

[2] R-squared measures the proportion of the wage variance explained by the regression's independent variables. In all regressions, the dependent variable is the average (across job incumbents) of the natural logarithm of hourly wages in the sampled job. Hence the R-squared statistics presented do not include the wage variation within sampled jobs, which is small relative to that across sampled jobs.

[3] The classification uses 42 occupational categories.

[4] More precisely, the estimated differential is exp (.168)-1 = .183, or 18.3 percent. Wage regression coefficients are reported throughout the text and then interpreted as approximate differentials. This practice is less precise but facilitates references to the tables.

[5] These surveys are restricted to establishments of at least 50 workers. Therefore, the estimated establishment employment coefficient does not reflect the wage-employment relationship in small establishments.

[6] For instance, to the extent that generic leveling data capture worker abilities, the differences in the regression coefficients in table 5 suggest that observed wage differences by establishment size or full-time status do partly reflect worker abilities, whereas wage differences due to union status do not.

[7] "Otherwise comparable" here means controlling for industry, occupation, establishment employment size, generic leveling factors, and all of the other covariates listed in table 5.

[8] As indicated in footnote 4, the precise premium is calculated as exp (.171) / exp (.089) minus 1, which equals .085, or 8.5 percent.

[9] More precisely, estimated wages in this occupation are exp (.62) which is 1.86 times as large as for the average occupation after controlling for other job and establishment attributes. Premiums in table 7 are the log wage regression coefficients from table 5 normed to be relative to the average occupational premium (instead of an omitted occupational category).

[10] Note that the regression pools different occupations, so the estimated returns to knowledge and other factors are not allowed to vary by occupation.

[11] The Office of Personnel Management originally designed generic leveling factors for the purposes of pay setting for white-collar Federal employees.