Compensating Differences and Interregional Wage Differentials
Author(s): Shelby D. Gerking and William N. Weirick
Source: *The Review of Economics and Statistics*, Vol. 65, No. 3 (Aug., 1983), pp. 483-487
Published by: The MIT Press
Stable URL: http://www.jstor.org/stable/1924194
Accessed: 26/05/2011 03:25

COMPENSATING DIFFERENCES AND INTERREGIONAL WAGE DIFFERENTIALS

Shelby D. Gerking and William N. Weirick*

I. Introduction

Interregional differences in average wages and earnings have been observed particularly in the North and South of the United States ever since the mid-1800s. That observation has motivated several empirical attempts to determine the source of those differentials, measured both in nominal and real terms, and to explain why they have been maintained over time. The general conclusion reached by the overwhelming majority of these studies is that the labor market has not eliminated these wage differentials even in the face of substantial interregional migration. This result has at least two alternative interpretations. First, it would appear to contradict the theory of compensating differences as applied to the labor market (Thaler and Rosen, 1975), which stresses that under the assumptions of perfect information, free geographic and intersectoral labor mobility, and homogeneous consumer tastes, the nominal wage rates of workers who have similar human capital characteristics, live and work in similar environments and experience similar living costs, are driven to equality. Second, this result may only reflect an aggregation error. In other words, there may be several types of labor that are each paid different equilibrium wage rates and comprise different percentages of the workforce in each region. Even if the real wage paid to each class of workers is interregionally invariant, a situation that instead would support the theory of compensating differences, failure to distinguish accurately between labor types could produce the illusion of a wage differential.

This paper considers the two alternative interpretations given above as to why interregional wage differentials might exist. Hedonic real wage equations are estimated for four regions of the United States using observations on individual household heads drawn from the 1976 Panel Study in Income Dynamics (PSID). This sample is of interest because the 1976 PSID data contain unusually detailed measures of education, work experience and occupation, as well as information on workplace and job characteristics. Thus, a more complete specification of the wage equation is permitted and the possibility of aggregation error is reduced, particularly in comparison with other interregional wage differential studies. Several of these studies, for example, have been based on aggregate data from the *Census of Manufactures* (Fuchs and Perlman, 1960; Gallaway, 1963; Scully, 1969; and Coelho and Ghali, 1971) which provide no direct measurements on the human capital of workers.

The remainder of the discussion is organized into three sections. Section II specifies the wage equation and describes the PSID data. Section III, then, reports empirical results which are consistent with the findings, based on aggregate data, of Bellante (1979) and Coelho and Ghali (1971) in that they support the theory of compensating differences. More specifically, for full-time workers, the rewards to attributes relevant in determining real wages apparently are interregionally invariant. However, because this result conflicts with most previous research on interregional wage differentials based on aggregate data and virtually all such research based on microdata (Welch, 1966; Hanoch, 1967; Hanushek, 1973, 1981; Hirsch, 1978; and Sahling and Smith, 1983), a number of empirical comparisons are made between the present study and the approaches taken by other investigators. Conclusions and implications are drawn out in section IV.

II. Specification of the Wage Equation

The general form of the hedonic wage equation considered is

$$WAGE = f(H, P, W, C) \qquad (1)$$

where $WAGE$ denotes the real wage paid, H denotes a vector of human capital characteristics, P denotes a vector of personal characteristics, W denotes a vector of work environment characteristics, and C denotes a vector of city attribute variables. Equation (1) is a reduced form which shows how, under the previously stated assumptions, both employers and workers have implicitly agreed to value the components of H, P, W, and C. If those assumptions hold at least approximately and if the United States as a whole is a relevant geographic perspective from which to consider the labor market, then the function f would be interregionally invariant. In other words, the hedonic prices associated with the components of H, P, W, and C would be identical across regions.

Received for publication April 26, 1982. Revision accepted for publication November 5, 1982.

* University of Wyoming, Laramie, and Northeast Louisiana University, Monroe, respectively.

We thank William D. Schulze, Sharon P. Smith, V. Kerry Smith, Anthony M. J. Yezer, and two anonymous referees for constructive comments on previous drafts of this paper and acknowledge partial financial support from the Environmental Protection Agency under Assistance Agreement CR 808893010. All views expressed in this paper are exclusively those of the authors and should not be attributed to officials at EPA. The usual caveat regarding responsibility for remaining errors applies here without qualification.

To more explicitly specify equation (1), the vector H contains measurements on household heads pertaining to: (1) years of full-time work experience, (2) months worked for present employer, (3) years of formal schooling, (4) advanced educational degrees, and (5) other schooling. The vector P contains measures of the household heads': (1) race, (2) sex, and (3) physical limitations and disabilities. The vector W contains measurements describing: (1) months required to become fully trained on present job, (2) the number of persons supervised, (3) union membership, (4) on-the-job injury rate in the industry where employed, and (5) occupation. The vector C, then, measures: (1) the size of the city in which the individual lives, (2) January temperature in that city, (3) local government expenditures per capita, (4) rate of reported crimes, (5) average annual precipitation, (6) average January windspeed. Exact descriptions and sources for these data are available from the authors on request.

A few selected comments on these variables are warranted. As indicated in the introduction, the 1976 PSID data contain unusually detailed information on human capital and workplace characteristics. For example, a direct measure of an individual's full-time work experience is provided; thus, the frequently used approximation, years of age minus years of schooling minus six, which overstates work experience levels for the intermittently unemployed, need not be used. Also, knowledge of how long the individual has worked for this present employer captures a related dimension of work experience, and the variable measuring the number of months required to become fully trained indicates the level of skills specific to the present job. A useful discussion of the role of these variables and other measures of training in determining wages is contained in a paper by Duncan and Hoffman (1979) who also analyzed data drawn from the 1976 PSID survey.

The dependent variable in equation (1) was defined as the real wage rate since the nominal wage paid, in equilibrium, should reflect living cost differences between geographic areas (Coelho and Ghali, 1971). The numerator of the real wage was described in the PSID survey with two variables, reported hourly wages received on the head's primary job, for those working for wages, and an hourly wage equivalent for those paid a salary. These two measures simply were merged into a single variable reflecting hourly compensation. Data on extra jobs and overtime pay were not utilized in this study. Nominal wages were then deflated by the Bureau of Labor Statistics (BLS) low budget, four-person cost of living index for Autumn 1975.[1] However, an obvious problem with this procedure is that the BLS computes the index only for certain SMSAs. Consequently, for the cases of PSID households not living in the counties where those SMSAs are located, the cost of living index was (somewhat arbitrarily) assigned the value assumed in the nearest BLS city.

For the purpose of estimating equation (1), the data set was reduced from the roughly 3,300 observations available to 1,741 after excluding all households where the head: (1) received more than 10% income from bonuses, commissions, overtime pay, and/or transfer payments, (2) worked less than 1,400 hours during 1975 and (3) was self-employed. The first of these exclusions was made in order to reduce the statistical problems created by families facing nonconvex budget constraints; the second was made in order to eliminate part-time workers from the sample, and the third was made so as to exclude those who may not be able to estimate accurately their annual hours of work.

III. Empirical Results

Ordinary least squares estimates of equation (1) were obtained after defining the dependent variable as the natural logarithm of the real wage and, in addition to the regressors already listed, including the squares of the variables measuring: (1) years of full-time experience, (2) months worked for present employer, (3) months required to become fully trained, (4) number of persons supervised and (5) years of schooling.[2] Actually, five separate identically specified regressions were run: a pooled regression using the entire 1,741 observation data set and four corresponding regressions based upon regional subsamples.[3] In all five regressions, the variables in H, P, and W generally are highly significant, plausibly signed, and at least roughly consistent with the results of previous empirical work.[4] However, the city attribute variables tended to perform less well with

[1] The 1976 PSID survey collected wage data from the previous year. Also, the low-budget cost of living index was used because the sample showed a slight tendency to be skewed toward the lower tail of the income distribution.

[2] Another regression was run with the dependent variable defined as the natural logarithm of the *nominal* wage, the natural logarithm of the price index as an explanatory variable, and all other explanatory variables specified as described earlier. The coefficient on the price index term was 0.97 which is not significantly different from unity at the 1% level.

[3] South includes the states of Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia. The West is composed of Arizona, California, Colorado, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming, Alaska, Hawaii. The following constitute the Northeast: Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont. The remainder define the North Central region: Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin.

[4] Results from these and all other regressions reported here are available from the authors on request.

TABLE 1.—TESTS FOR INTERREGIONAL DIFFERENCES IN EQUATION STRUCTURE (1400 HOURS)

Regression	Real Wage	Nominal Wage	Sahling-Smith	Real Earnings	Nominal Earnings
		Error Sum of Squares			
Pooled (1,741)[a]	146.17	147.42	190.07	223.66	225.15
Northeast (350)[a]	28.35	28.11	32.54	35.04	34.92
North Central (402)[a]	27.18	26.77	39.16	44.30	43.76
South (654)[a]	54.08	54.95	73.52	90.81	92.06
West (335)[a]	26.91	27.02	29.34	38.76	38.76
			F-Statistics		
No. Exp. Var.	32	32	29	32	32
$F(96,1613)$	1.19	1.30[b]	—	1.19	1.26[b]
$F(87,1625)$	—	—	1.66[c]	—	—

[a] Number of observations shown in parentheses.
[b] Denotes significantly greater than unity at 5% level.
[c] Denotes significantly greater than unity at 1% level.

only the two dummies reflecting city size turning out to be statistically significant at the 5% level.

These estimated wage equations form the basis for a Chow (1960) test of the null hypothesis that the true intercepts *and* slopes are identical between regions. The result of that test, which is shown in the second column of table 1 labelled "real wage" is the statistic $F(96,1613) = 1.188$, which is not significantly greater than unity at the 5% level.[5] Also, after dropping the 7 regressors which were never significantly different from zero at the 5% level either in the pooled sample or regional regressions, the real wage equation was re-estimated using the pooled sample with 3 dummy variables to allow for regional intercept shifts and 72 interaction variables (3 for each of the 24 remaining regressors excluding the constant term) to allow for regional slope shifts.[6] Only 3 of the estimated coefficients of those additional 75 variables were significantly different from zero.[7] Therefore, after adjusting for cost of living differences, workers in the sample with similar human capital, personal, work environment, and city attribute characteristics have approximately the same pay rates no matter where they live. That result is consistent with Rosen's theory of hedonic price determination in implicit markets, and supports Bellante's contention that differences in endowments of various heterogeneous labor types are responsible for a large share of the observed interregional differences in average real wages. Moreover, it is in a sense parallel to Duncan and Hoffman's (1979) finding that the payoff of on-the-job training is independent of race and sex.

As indicated in the introduction, however, the results reported conflict with the findings of the majority of investigators who have examined interregional variations in the structure of wages and earnings. Three factors are hypothesized to be at least partially responsible: (1) the treatment of cost of living differences between regions, (2) the completeness of the specification of the wage or earnings equation, and (3) the treatment of hours worked.

A. *Cost of Living Differences*

The variation in cost of living differences between geographic areas may be important to consider since, according to the theory of compensating differences, workers would require a wage premium in order to induce them to remain in an area where living costs are comparatively high. Nevertheless, despite the arguments made by Coelho and Ghali (1971), in many analyses of wages, income, and earnings, including those by Griliches and Mason (1972), Hanushek (1973, 1981), Hirsch (1978), and Brown (1980), no explicit account is taken of that variable. Neglecting cost of living differences in the PSID data set does lead to a finding of significant interregional variation in wage equation structure, although numerically the appropriate F-statistic is not much higher than the 1.188 figure reported earlier. More specifically, the pooled and regional regressions just described were re-estimated using the natural logarithm of the *nominal* wage as the dependent variable and with all regressors unchanged. As shown in

[5] A parallel Chow test was performed after re-estimating each of the five equations using only the 945 observations drawn for heavily populated counties for which there was no ambiguity in the assignment of the cost of living index. The resulting F-statistic was 1.06 which is also not significantly greater than unity at the 5% level.

[6] The seven omitted regressors were those measuring temperature, windspeed, local government expenditures, crime, and precipitation, as well as dummy variables for the occupations of laborer and sales worker.

[7] The three variables on which coefficients were significantly different from zero at the 5% level were (1) the product of advanced educational degrees and West, (2) the product of physical limitations and disabilities and South, and (3) the product of sex and North Central.

column 3 of table 1, the resulting Chow test F-statistic increased slightly to 1.298, which is barely significantly greater than unity at the 5% level.

B. *Equation Specification*

A second possible explanation for why the present study finds no significant interregional variations in wage equation structure is that the PSID data allow for a more fully specified set of human capital and work environment variables than is possible with other data sets. For example, Sahling and Smith (1983) divided Current Population Survey (CPS) data into five regions (metropolitan New York, Northeast, North Central, South, and West) and, for their pooled sample as well as for each region, real wages were regressed on: (1) years of schooling, (2) square of years of schooling, (3) years of work experience (measured as years of age minus years of schooling minus six), (4) square of years of work experience, (5) the product of experience and schooling, (6) marital status, (7) race, (8) Spanish origin, (9) veteran status (for males), (10) regular part-time worker, (11) dual job holder (12) union member, (13) occupation, and (14) industry of employment. Statistical tests showed significantly different wage equation structures between the five regions. To further analyze that conclusion, first note that the CPS data contain no measures of key variables such as months worked for present employer, months required to become fully trained on present job, and the number of persons supervised. Also, work experience is measured only indirectly. Additionally, using the 1,741 observation PSID data set, the pooled sample and regional real wage equations were re-estimated with the regressors specified similarly to those used by Sahling and Smith. The resulting Chow test, reported in column 4 of table 1, produced the statistic $F(87,1625) = 1.888$, which is significantly different from unity at the 1% level.

C. *Hours Worked*

A third difference between the present study and previous analyses of interregional wage differentials lies in the treatment of hours worked. Hanoch (1967), Hanushek (1973, 1981), and Hirsch (1978), for example, use nominal labor earnings as the dependent variable rather than the wage. As evidenced by columns 5 and 6 of table 1, however, the Chow test F-statistics that result from using the natural logarithm of nominal and real earnings parallel those obtained when their wage counterparts are used (all regressors were defined identically in the four sets of equations). When the dependent variable was defined as real earnings, the statistic $F(96,1613) = 1.186$ was obtained, which is not significantly greater than unity at the 5% level, whereas, when the real earnings variable was replaced by the natural logarithm of nominal earnings, the corresponding F-statistic rose slightly in numerical magnitude, but by enough to make it significantly different from unity at the 5% level.

TABLE 2.—TESTS FOR INTERREGIONAL DIFFERENCES IN EQUATION STRUCTURE (520 HOURS)

Regression	Error Sum of Squares		
	Real Wage	Sahling-Smith	Real Earnings
Pooled (1,984)[a]	195.51	226.12	410.45
Northeast (393)[a]	36.39	41.01	58.76
North Central (456)[a]	36.59	49.88	84.72
South (742)[a]	66.18	83.39	170.04
West (393)[a]	35.26	36.27	66.98
		F-Statistics	
No. Exp. Var.	32	29	32
$F(96,1855)$	2.34[b]	—	1.52[b]
$F(87,1868)$	—	1.59[b]	—

[a] Number of observations shown in parentheses.
[b] Denotes significantly greater than unity at 1% level.

Also, other investigators including Sahling and Smith (1983) use a smaller number of annual hours worked as the lower bound for including workers in their sample as compared with the 1,400 annual hours used here. That alteration appears to have a relatively greater effect on the results, a situation demonstrated in table 2. Column 2 of table 2 indicates that when the "real wage" regressions reported previously were rerun, the Chow test F-statistic more than doubled to 2.34 (compared with column 2 of table 1). Moreover, setting minimum annual hours equal to 520 and either using the natural logarithm of real earnings as the dependent variable or adopting the Sahling and Smith specification produces an F-statistic that exceeds unity at the 1% level. Therefore, the conclusion of interregional invariance of wage equation structure would appear to apply more directly to full-time as compared with part-time workers. That result should not be surprising since part-time workers are more likely to be tied to a particular geographic area for nonemployment related reasons. Additionally, the smaller the number of hours worked, the lower would be the incentive to move in response to an interregional wage differential of a given size.

IV. Summary and Conclusions

This paper has presented evidence, based on microdata from the 1976 Panel Study in Income Dynamics, concerning the nature of interregional wage differentials in the United States. The results presented, which are consistent with the theory of compensating differences in the labor market, support the hypothesis that a full-time worker's real wages or earnings do not differ between broadly defined geographic areas of the United

States. Rather, observed interregional differences in *average* real wages probably arise from different relative endowments of various heterogeneous labor types. Because these results conflict with findings of most previous studies, comparisons are made with the approaches taken by other investigators. Those comparisons indicate that empirical estimates of interregional differences in the structure of wage and earnings equations are sensitive to (1) the treatment of geographic cost of living differences, (2) the completeness of the specification of the regressors, particularly the human capital measures, and (3) whether part-time workers are included in the sample.

REFERENCES

Bellante, Don, "The North-South Differential and the Migration of Heterogeneous Labor," *American Economic Review* 69 (Mar. 1979), 166–175.

Brown, Charles, "Equalizing Differences in the Labor Market," *Quarterly Journal of Economics* 96 (Feb. 1980), 113–134.

Chow, Gregory, "Tests of Equality between Subsets of Coefficients in Two Linear Regressions," *Econometrica* 28 (July 1960), 591–605.

Coelho, Philip, and Moheb Ghali, "The End of the North-South Wage Differential," *American Economic Review* 61 (Dec. 1971), 932–938.

Duncan, Greg, and Saul Hoffman, "On-the-Job Training and Earnings Differences by Race and Sex," this REVIEW 61 (Nov. 1979), 594–603.

Fuchs, Victor, and Richard Perlman, "Recent Trends in Southern Wage Differentials," this REVIEW 42 (Aug. 1960), 292–300.

Gallaway, Lowell, "The North-South Wage Differential," this REVIEW 45 (Aug. 1963), 264–272.

Griliches, Zvi, and William Mason, "Education, Income, and Ability," *Journal of Political Economy* 80 (May/June 1972), S74-S103.

Hanoch, Giora, "An Economic Analysis of Earnings and Schooling," *Journal of Human Resources* 2 (Summer 1967), 310–329.

Hanushek, Eric, "Regional Differences in the Structure of Earnings," this REVIEW 55 (May 1973), 204–213.

———, "Alternative Models of Earnings Determination and Labor Market Structures," *Journal of Human Resources* 26 (Spring 1981), 238–259.

Hirsch, Barry, "Predicting Earnings Distributions across Cities: The Human Capital Model vs. the National Distribution Hypothesis," *Journal of Human Resources* 13 (Summer 1978), 336–384.

Sahling, Leonard, and Sharon Smith, "Regional Wage Differentials: Has the South Risen Again?" this REVIEW 65 (Feb. 1983), 131–135.

Scully, Gerald, "Interstate Wage Differentials: A Cross-Section Analysis," *American Economic Review* 59 (Dec. 1969), 757–773.

Thaler, Richard, and Sherwin Rosen, "The Value of Saving a Life: Evidence from the Labor Market," in N. Terleckyj, (ed.), *Household Production and Consumption* (New York: NBER, 1975).

Welch, Finis, "Measurement of the Quality of Schooling," *American Economic Review* 56 (May 1966), 379–392.

DISAGGREGATION AND THE LABOR PRODUCTIVITY INDEX

Jack H. Beebe and Jane Haltmaier*

Where input and output data are aggregated directly across sectors as in the National Income and Product Accounts (NIPA), intersectoral shifts impact on measured aggregate productivity change. Although most productivity studies use crude methods to approximate the effects of intersectoral shifts, some authors have sought to derive exact formulae for measuring the shift effect. In a different direction, others have employed Divisia aggregation, which gives a measure of aggregate productivity change that is net of intersectoral shift effects.

This paper derives a simple and exact formula for decomposing aggregate productivity change into "rate," "level," and "interaction effects" and applies this formula to post-WWII secular labor-productivity data using both 12-sector and 60-sector disaggregation. The literature is first reviewed briefly and the rate, level, and interaction effects derived algebraically. Then the formulae are applied to secular trends in labor productivity over the 1948–78 period using commonly accepted subperiods of peak-to-peak labor productivity performance —1948–65, 1965–73, and 1973–78.

Received for publication March 26, 1981. Revision accepted for publication September 24, 1982.

* Federal Reserve Bank of San Francisco, and Board of Governors of the Federal Reserve System, respectively.

We thank Tom Klitgaard of Stanford University for his excellent research assistance.

Algebraic Formulation

Published labor productivity data are calculated using direct aggregation: outputs are added across sectors, labor inputs are also summed, and total output is then divided by total labor input to arrive at a calculated aggregate level of average labor productivity. Aggregate average labor productivity is affected over time by productivity change within each sector and by shifts of output and employment among sectors with different levels of average productivity. In netting out interindustry shift effects, some authors (notably, Denison (1979) and Kendrick (1980)) correct for labor shifts