Volume Title: International and Interarea Comparisons of Income, Output, and Prices

Volume Author/Editor: Alan Heston and Robert E. Lipsey, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-33110-5

Volume URL: http://www.nber.org/books/hest99-1

Publication Date: January 1999

Chapter Title: Constructing Interarea Compensation Cost Indexes with Data from Multiple Surveys

Chapter Author: W. Brooks Pierce, John W. Ruser, Kimberly D. Zieschang

Chapter URL: http://www.nber.org/chapters/c8388

Chapter pages in book: (p. 171 - 210)

4 Constructing Interarea Compensation Cost Indexes with Data from Multiple Surveys

W. Brooks Pierce, John W. Ruser, and Kimberly D. Zieschang

Place-to-place compensation cost comparisons for areas within the United States are very much in demand to inform facility location decisions and locality salary administration policy in both the private and the public sectors. The Bureau of Labor Statistics (BLS) has operated geographically comprehensive surveys measuring locality wage levels since 1991, primarily to support the Federal Employee Pay Comparability Act (FEPCA). This law was enacted to align the locality compensation for federal employees with that of the comparable nonfederal workforce. Similar studies oriented more toward place-to-place comparisons of compensation are undertaken in the private sector by several companies, often with specialties in certain industry and/or occupation groups.

Labor services differ in type and quality from area to area. The central problem this paper considers is how indexes comparing employee compensation costs across geographic areas that account for the heterogeneity of jobs and workers may be formulated and calculated. A long-standing approach to the heterogeneity problem, taken, for example, by the BLS in the Occupational Compensation Survey program, is to make interarea comparisons only between the same narrowly defined jobs that exist in every area for which comparisons are to be made. The limitation of this approach is that the comparisons apply only to the population of jobs that are found in all areas, while jobs that are specific to only certain areas are excluded from the comparisons.

W. Brooks Pierce is a senior economist and John W. Ruser the chief economist in the Compensation Research and Program Development Group of the Bureau of Labor Statistics' Office of Compensation and Working Conditions. Kimberly D. Zieschang is a senior economist with the International Monetary Fund. This paper was completed while he was associate commissioner for compensation and working conditions at the Bureau of Labor Statistics, U.S. Department of Labor.

The views expressed are those of the authors and do not necessarily reflect the views or policies of the Bureau of Labor Statistics or the International Monetary Fund.

Another approach, which we follow in this paper, is to define jobs more broadly by industry and occupation and to utilize data for all jobs to make interarea comparisons. Within these broader groups, the compensation rates received by specific jobs are then related to specific quantitative information on the characteristics of the jobs using a statistical model known in the economics and economic statistics literature as a *hedonic* model. Heterogeneity is controlled for by employing the parameters estimated in these hedonic models, fitted using regression analysis, to adjust for observable characteristics of workers and jobs. This approach has the advantage of covering all jobs in each labor market. However, it requires additional information about jobs or workers that can be used as covariates in the hedonic regressions.

To provide a context and rigorous interpretation for our indexes, we begin with a standard microeconomic framework for input price indexes developed in a long economic index number literature, positing a model of producer input cost minimization conditioning on output and exogenously determined input prices. A good statement of this basic economic input price index framework applied to labor input cost measurement is given in Triplett (1983). Triplett also discusses the application of hedonic regression methods to adjust for labor quality.

We take these index number concepts and hedonic labor quality measurement methods and incorporate them into an integrated, computable index number system. To construct place-to-place compensation comparisons, we use a Törnqvist index formula. We adopt the Törnqvist formula, rather than alternatives also used in geographic price comparisons such as the Geary (1958)–Khamis (1970) "international prices" system or an adjusted Fisher ideal approach, because the Törnqvist framework simultaneously displays five important features:

First, the Törnqvist formulas that we use for bilateral comparisons have been shown by Diewert (1976) to be exact for the translog flexible functional form. By implication, they accurately accommodate producers' substitution decisions among types of labor services as their relative prices differ from place to place.

Second, Caves, Christensen, and Diewert (1982a) have shown that the Törnqvist index number is exact even when there are significant differences in the underlying technology between situations compared. These indexes therefore accommodate variations in the technology that producers select as they consider various site locations.

Third, Kokoski, Moulton, and Zieschang (chap. 3 in this volume) provide a closed form for the class of all-systems Törnqvist bilateral index numbers that are transitive, generalizing a result along these lines introduced by Caves, Christensen, and Diewert (1982b), and provide a feasible algorithm for computing such systems of index numbers. Clearly, use of our compensation indexes to inform a salary administration policy for geographically dispersed organizations would require the transitivity property to eliminate the possibility of gains and losses accruing to reassigned staff as a sole result of a series of relocations.

Fourth, Kokoski, Moulton, and Zieschang (chap. 3 in this volume) demonstrate that the Törnqvist multilateral system of parities will aggregate in a natural way with respect to a given classification hierarchy for types of labor. This facilitates explaining variations in aggregate compensation in terms of variations in the component occupations making up the aggregate, an important property for a system of public compensation statistics.

Fifth, Kokoski, Moulton, and Zieschang (chap. 3 in this volume) have adapted earlier exact index number results from Zieschang (1985, 1988) and Fixler and Zieschang (1992a, 1992b) to incorporate information on variation in the characteristics of detailed items when making Törnqvist index number comparisons. In this index number framework, coefficients from hedonic compensation regressions are used in constructing labor quality–adjustment factors for place-to-place indexes of compensation rates. Because of the heterogeneity in the measured characteristics of labor employed within industry/occupation groups across areas, these exact quality adjustments are important for making accurate compensation comparisons from place to place.

In section 4.1, we briefly state the microeconomic foundations of our approach using standard production theory and show how the aggregate conceptual bilateral area indexes of this framework can be operationalized using Törnqvist exact and superlative index numbers. We then show how the differing measured labor services characteristics that are encountered in various areas can be accounted for in the index framework. In section 4.2, we establish the implications of transitivity in our Törnqvist interarea index system, and, in section 4.3, we show how transitivity can be imposed with minimal adjustment of the data. In section 4.4, we apply this methodology to both establishment micro data on jobs from the U.S. Employment Cost Index (ECI) and area/occupation/industry data on workers from the Current Population Survey (CPS) and present labor services characteristics–adjusted interarea wage and compensation indexes for thirty-nine major urban centers and rest-of-regional-division geographic areas. We conclude in section 4.5.

4.1 Economic Index Number Concepts Incorporating Information on the Characteristics of Heterogeneous Labor Services

Let p_i^a be the price or compensation rate in area a, of which there are A areas in total, of labor services of occupation i. Let q_i^a be the corresponding quantity purchased, and let x_i^a be the vector of characteristics of the ith job specification for labor services transacted in area a. Let e_h^a represent the total labor services expense of establishment h in area a, and let q_h^a denote the vector of labor services consumed by establishment h in area a with vector of characteristics x_h^a and prices p_h^a.

We suppose that each establishment in area a minimizes the cost of achieving a given level of output u_h^a at expenditure level e_h^a so that the establishment expense incurred for a given quality of labor services as determined by the vector x_h^a would be

$$e_h^a = E_h^a(u_h^a, x_h^a, p_h^a) = \min_{q_h^a}\{p_h^{a\prime} q_h^a : d_h^a(u_h^a, x_h^a, q_h^a) \geq 1\},$$

where $d_h^a(u_h^a, x_h^a, q_h^a)$ is the joint production function of establishment h.[1] To reduce clutter, we condition on and suppress the nonlabor inputs used by establishment h.

We suppose further that an establishment in area a faces a hedonic locus of market equilibrium prices across the labor services quality spectrum given by $p_h^a = H^a(x_h^a)$ and that the establishment minimizes the cost of achieving outputs u_h^a over the characteristics of labor services, with the result that

$$\text{(1)} \qquad \nabla_{x_h^a} E_h^a(u_h^a, x_h^a, p_h^a) + \nabla_{x_h^a} p_h^{a\prime} \nabla_{p_h^a} E_h^a(u_h^a, x_h^a, p_h^a) = 0.$$

Since $\nabla_{x_h^a} p_h^a = \nabla_{x_h^a} H^a$ and $\nabla_{p_h^a} E_h^a(u_h^a, x_h^a, p_h^a) = q_h^a$, the latter by the Shephard/Hotelling lemma, we have

$$\text{(2)} \qquad \nabla_{x_h^a} E_h^a(u_h^a, x_h^a, p_h^a) = -\nabla_{x_h^a} H^{a\prime} q_h^a.$$

If H^a is semilog, as generally assumed in hedonic studies, so that

$$\text{(3)} \qquad \ln H_i^a = \alpha_i^a + \beta_i^{a\prime} x_i^a,$$

then the characteristics gradient expression can be rewritten

$$\text{(4)} \qquad \nabla_{x_h^a} E_h^a(u_h^a, x_h^a, p_h^a) = -\beta^{a\prime} w_h^a e_h^a,$$

where

$$w_{i,h}^a = \frac{p_{i,h}^a q_{i,h}^a}{\sum_i p_{i,h}^a q_{i,h}^a},$$

$$w_h^a = \begin{bmatrix} w_{1,h}^a \\ \vdots \\ w_{N_q,h}^a \end{bmatrix}; \quad N_q = \text{number of types (occupations) of labor, and}$$

$$\beta^{a\prime} = \begin{bmatrix} \beta_1^{a\prime} \\ \vdots \\ \beta_{N_x}^{a\prime} \end{bmatrix}; \quad N_x = \text{number of labor services characteristics.}$$

1. This particular joint production function is the *input distance function,* given by

$$d_h^a(u_h^a, x_h^a, q_h^a) = \sup_\theta \left\{\theta : \left(u_h^a, x_h^a, \frac{1}{\theta} q_h^a\right) \text{is feasible}\right\}.$$

Diewert (1987) considers the area aggregation of individual establishments in the context of an area production function. We follow this general notion but will require some modifications to handle the heterogeneity of labor service types and their prices within and between areas. Turning now to aggregate labor input expense over establishments in an area, we have

$$E^a(\vec{u}^a, \vec{x}^a, \vec{p}^a) = \sum_h E_h^a(u_h^a, x_h^a, p_h^a),$$

where the arrow over an argument indicates the concatenation of vectors across establishments. We then consider the labor services expenditure or cost function in terms of log transformed price arguments as

$$\begin{aligned} Q^a(\vec{u}^a, \vec{x}^a, \ln\vec{p}^a) &= E^a(\vec{u}^a, \vec{x}^a, \vec{p}^a) \\ &= \sum_h E_h^a(u_h^a, x_h^a, p_h^a) = \sum_h Q_h^a(u_h^a, x_h^a, \ln p_h^a). \end{aligned}$$

We aggregate across establishments in area a such that the expenditure-weighted average for characteristics and log-labor services prices represents the indicators determining area demand behavior, where area item demand for labor services is the sum of the establishment item demands for the area. We do not require strong aggregation conditions but effectively hold the distribution of labor services characteristics and compensation rates fixed across establishments within area a as in

$$\text{(5)} \quad \tilde{Q}^a\left(\vec{u}^a, \bar{x}^a, \overline{\ln p}^a\right) = Q^a\left(\vec{u}^a, \iota \otimes \bar{x}^a + v_x^a, \iota \otimes \overline{\ln p}^a + v_{\ln p}^a\right),$$

where $v_x^a = \vec{x}^a - \iota \otimes \bar{x}^a$, $v_{\ln p}^a = \ln\vec{p}^a - \iota \otimes \overline{\ln p}^a$, ι = a vector of ones equal in dimension to the number of establishments in area a, and $\otimes$ = Kronecker product, which give the deviations of the area means from the individual establishment values for labor services characteristics and log compensation rates paid.

Using the derivatives of the expenditure function with respect to log prices expressed in terms of observable expenditure shares, Diewert (1976) and Caves, Christensen, and Diewert (1982a) have shown that the Törnqvist index number is exact for the translog flexible functional form, which differentially approximates any price aggregator function (i.e., cost of utility, input cost, revenue function) to the second order at a point, and it is exact even when some of the parameters (those on the first-order terms) of the underlying aggregator function are different in the two periods or localities compared. We take the derivative of the area expenditure function with respect to establishment labor cost–weighted aggregate arguments to obtain

$$\text{(6)} \quad \begin{aligned} \frac{\partial}{\partial \bar{x}_{iz}^a} \ln \tilde{E}^a\left(\vec{u}^a, \bar{x}^a, \exp(\overline{\ln p}^a)\right) &= \frac{\partial}{\partial \bar{x}_{iz}^a} \ln \tilde{Q}^a\left(\vec{u}^a, \bar{x}^a, \overline{\ln p}^a\right) \\ &= \sum_h \frac{\partial}{\partial x_{izh}^a} Q_h^a\left(u_h^a, x_h^a, \ln p_h^a\right)/\tilde{Q}^a = -\beta_{izx}^a \sum_H w_{ih}^a s_h^a = \beta_{izx}^a \overline{w}_i^a, \end{aligned}$$

$$\begin{aligned}(7)\quad &\frac{\partial}{\partial \ln p_i^a} \ln \tilde{E}(\vec{u}^a, \bar{x}^a, \exp(\overline{\ln p^a})) = \frac{\partial}{\partial \ln p_i^a} \ln \tilde{Q}(\vec{u}^a, \bar{x}^a, \overline{\ln p^a}) \\ &= \sum_h \frac{\partial}{\partial \ln p_{izh}^a} Q_h^a(u_h^a, x_h^a, \ln p_h^a)/\tilde{Q}^a = \sum_h w_{ih}^a s_h^a = \overline{w}_i^a,\end{aligned}$$

where

$$w_{ih}^a = \frac{p_{ih}^a q_{ih}^a}{\sum_i p_{ih}^a q_{ih}^a} = \frac{p_{ih}^a q_{ih}^a}{e_h^a},$$

$$s_h^a = \frac{e_h^a}{\sum_h e_h^a},$$

are, respectively, the within-firm labor cost shares of occupations and the between-firm labor cost shares of establishments in area a.

Finally, we assume that the area aggregate labor services cost function $\ln \tilde{Q}^a (u^a, \bar{x}^a, \overline{\ln p^a})$ has a quadratic, "semi-translog" functional form in its arguments with coefficients of second-order terms independent of location but with possibly location-specific coefficients on linear terms. Following Caves, Christensen, and Diewert (1982a), then, we can derive the following (logarithmic) index number result:

$$\begin{aligned}&\ln I^{ab} \\ &= \frac{1}{2}[\ln \tilde{Q}^a(\vec{u}^a, \bar{x}^b, \overline{\ln p^b}) - \ln \tilde{Q}^a(\vec{u}^a, \bar{x}^a, \overline{\ln p^a}) \\ (8)\quad &+ \ln \tilde{Q}^b(\vec{u}^b, \bar{x}^b, \overline{\ln p^b}) - \ln \tilde{Q}^b(\vec{u}^b, \bar{x}^a, \overline{\ln p^a})] \\ &= \frac{1}{2}[\nabla_{\ln p} \ln \tilde{Q}^a(\vec{u}^a, \bar{x}^a, \overline{\ln p^a}) + \nabla_{\ln p} \ln \tilde{Q}^b(\vec{u}^b, \bar{x}^b, \overline{\ln p^b})](\overline{\ln p^b} - \overline{\ln p^a}) \\ &+ \frac{1}{2}[\nabla_x \ln \tilde{Q}^a(\vec{u}^a, \bar{x}^a, \overline{\ln p^a}) + \nabla_x \ln \tilde{Q}^b(\vec{u}^b, \bar{x}^b, \overline{\ln p^b})](\bar{x}^b - \bar{x}^a).\end{aligned}$$

Substituting (6) and (7) into (8), we have

$$\begin{aligned}(9)\quad \ln I^{ab} = \ln T^{ab} \equiv \frac{1}{2}\sum_i \Big[&(\overline{w}_i^a + \overline{w}_i^b)(\overline{\ln p_i^b} - \overline{\ln p_i^a}) \\ &- \sum_z (\beta_{iz}^a \overline{w}_i^a + \beta_{iz}^b \overline{w}_i^b)(\bar{x}_{iz}^b - \bar{x}_{iz}^a)\Big].\end{aligned}$$

This is an extremely flexible result that permits all parameters of the semilog "hedonic" labor services compensation equations to differ by area and reflects establishments' optimizing behavior in considering location and the available characteristics of labor services.

4.2 Törnqvist Multilateral (Transitive) Systems of Bilateral Compensation Index Numbers

In another paper, Caves, Christensen, and Diewert (1982b) noted that the system of bilateral Törnqvist interarea indexes is not transitive but developed a simply calculated multilateral variant satisfying the transitivity property. Following Kokoski, Moulton, and Zieschang (chap. 3 in this volume), we apply the following general implication of transitivity for this class of index number:

$$\begin{aligned}
&\sum_n w_i^a\left(\ln p_i^b - \left(\sum_z \beta_{iz}^a x_{iz}^b\right)\right) - \left[\sum_i w_i^a\left(\ln p_i^a - \left(\sum_z \beta_{iz}^b x_{iz}^a\right)\right)\right] \\
(10)\qquad &= \sum_i\sum_z[-\beta_{iz}^0 w_i^0](x_{iz}^b - x_{iz}^a) + \sum_i\sum_z x_{iz}^0(\beta_{iz}^b w_i^b - \beta_{iz}^a w_i^a) \\
&\quad + \sum_i w_i^0(\ln p_i^b - \ln p_i^a) + \sum_i[-\ln p_i^0](w_i^b - w_i^a),
\end{aligned}$$

where x_{iz}^0 = a reference characteristic z for index item i across the entire region, β_{iz}^0 = a reference coefficient for the characteristic z of item i in a semilog hedonic equation explaining specification price across the entire region, p_i^0 = a reference price for item i across the entire region, and w_i^0 = a reference share for item i for the entire region, where $\sum_i w_i^0 = 1$. If this condition holds, the multilateral Törnqvist index has the form

$$\begin{aligned}
\ln T^{ab} &= -\sum_i\sum_z \frac{1}{2}(\beta_{iz}^0 w_i^0 + \beta_{iz}^b w_i^b)(x_{iz}^b - x_{iz}^0) + \sum_i \frac{1}{2}(w_i^0 + w_i^b)(\ln p_i^b - \ln p_i^0) \\
(11)\qquad &-\left[-\sum_i\sum_z \frac{1}{2}(\beta_{iz}^0 w_i^0 + \beta_{iz}^a w_i^a)(x_{iz}^a - x_{iz}^0) + \sum_i \frac{1}{2}(w_i^0 + w_i^a)(\ln p_i^a - \ln p_i^0)\right].
\end{aligned}$$

The proof is given in Kokoski, Moulton, and Zieschang (chap. 3 in this volume). Caves, Christensen, and Diewert (1982b) showed that application of the EKS principle to a system of bilateral Törnqvist indexes yields the price component of the above formula with the reference shares and log prices set at their simple arithmetic averages across areas. Clearly, these simple averages could also be calculated as total compensation expenditure–weighted averages. Extension for the EKS/Caves, Christensen, and Diewert (CCD) approach to our labor quality–adjusted index given by equation (11) would simply require that the zero-superscripted terms constituting the product of the reference hedonic coefficients of each characteristic with the reference share weight of the index items be set to the regional averages for these terms. In this paper, however, we use the Kokoski, Moulton, and Zieschang regression method for determining the reference parameters, as detailed in section 4.3 below.

4.2.1 Analysis of the Contribution of Labor Quality Indicators to Levels of Place-to-Place Indexes

Because Törnqvist indexes are linear in the log differences of detailed, quality-adjusted specification prices, the contribution of each quality indicator, say, full-time status, to the quality-level ratio between two areas can be readily calculated by exponentiating the appropriate weighted sums of log price differences. These sums would be calculated from the transitive expression for the index given above, where it is expressed in terms of locality weights averaged with reference weights and price differentials from reference prices. The contribution to the level of ln T^{ab} of labor characteristic z would simply be the subordinate sum

$$\ln C_z^{ab} = -\sum_i \frac{1}{2}(\beta_{iz}^0 w_i^0 + \beta_{iz}^b w_i^b)(x_{iz}^b - x_{iz}^0)$$
$$- \left[-\sum_i \frac{1}{2}(\beta_{iz}^0 w_i^0 + \beta_{iz}^a w_i^a)(x_{iz}^a - x_{iz}^0)\right].$$

4.3 Estimation of the Reference Values for Shares, Prices, and Determinants of Quality

4.3.1 Adjusting for Labor Quality from Place to Place

In the present study, we utilize wage and compensation cost data from the employment cost index (ECI) survey. This survey contains a limited amount of information about each surveyed job. We augment the observed characteristics of jobs with additional data on worker characteristics from the Current Population Survey (CPS).

We follow the method of Kokoski, Cardiff, and Moulton (1994), who construct interarea price indexes for consumer goods using country-product-dummy (CPD) regression (Summers 1973). We first estimate wage and compensation costs regressions for each broadly defined job, where the covariates include worker and job attributes and local area dummies. Let p_{ij}^a represent the wage in the jth quote for job i in location a, where a job is defined to be in an industry/occupation group. The wage can be described by the following regression equation:

(12) $$\ln p_{ij}^a = X_{ij}^a \beta_i + L_i^a + \varepsilon_{ij}^a,$$

where X_{ij}^a represents data on the characteristics of the job and the worker and where L_i^a represents a local area effect for job i in area a. This regression equation allows the coefficients on X_{ij}^a and the local area effects to vary across jobs. Equation (12) is estimated by weighted least squares, where the weights are the sample weights from the ECI.

A standard practice is to utilize the estimation results from equation (12) to make interarea wage comparisons. The regression defines a decomposition of interarea wage differences into components due to interarea differences in attributes X_{ij}^a and residual terms L_i^a. Let $\hat{\beta}_{iz}$ be the zth element of the vector of weighted least squares estimates of β_i from equation (12), and let x_{ijz}^a be the zth element of the vector X_{ij}^a. Also let $\overline{\ln p_i^a}$ and $\bar{x}_{iz}^a$ be weighted (by ECI sample weights) averages over j of $\ln p_{ij}^a$ and x_{ijz}^a, respectively. Then, by the properties of the weighted least squares estimators,

$$\overline{\ln p_i^a} = \sum_z \hat{\beta}_{iz}\bar{x}_{iz}^a + \hat{L}_i^a.$$

A Törnqvist index comparing wages in local area b to those in area a is defined (in logs) as

$$\ln T^{ab} = \frac{1}{2}\sum_i \left(w_i^a + w_i^b\right)\left(\overline{\ln p_i^b} - \overline{\ln p_i^a}\right), \tag{13}$$

where w_i^a is cell i's share of the labor expenditure in locality a. This differential can in effect be decomposed into contributions of the various covariates in X and contributions of the local area dummies. The contribution of the local area dummies takes the same form as (13),

$$\ln T_L^{ab} = \frac{1}{2}\sum_i (w_i^a + w_i^b)(\hat{L}_i^b - \hat{L}_i^a).$$

Further, the contribution of the zth characteristic of the job or worker to the index in (13) is

$$-\frac{1}{2}\sum_i (w_i^a + w_i^b)\hat{\beta}_{iz}(\bar{x}_{iz}^b - \bar{x}_{iz}^a).$$

This contribution depends on interarea differences in average characteristics (the difference in the mean X's) in conjunction with the importance of the zth characteristic in determining the wages in each job i (the $\hat{\beta}_{iz}$). The sum of these z contributions, plus $\ln T_L^{ab}$, equals the Törnqvist index in (13). In the following sections, we present this decomposition for a (transitive, multilateral) set of Törnqvist bilateral index numbers.

4.3.2 Multilateral Compensation Indexes: A Regression Approach for Imposing Transitivity with Minimal Adjustment of the Data

In this paper, we employ an alternative to (or a likely superclass of) the EKS/CCD approach from Kokoski, Moulton, and Zieschang (chap. 3 in this volume) for making the system of bilateral indexes transitive. When this condition on the cross-weighted differences of labor characteristics–adjusted log regional prices is not met, the data may be minimally adjusted to satisfy transitivity by fitting the equation

$$w_i^a\left(\ln p_i^b - \left(\sum_z \beta_{iz} x_{iz}^b\right)\right) - w_i^a\left(\ln p_i^a - \left(\sum_z \hat{\beta}_{iz} x_{iz}^a\right)\right)$$
$$= [w_i^0]\left(\ln p_i^b - \sum_z \hat{\beta}_{iz} x_{iz}^b - \left(\ln p_i^a - \sum_z \hat{\beta}_{iz} x_{iz}^a\right)\right)$$
$$+ [-\ln p_i^0](w_i^b - w_i^a) + \varepsilon_i^{ab}$$

using least squares to obtain the estimates

$$\begin{bmatrix} -\ln \hat{p}_i^0 \\ \hat{w}_i^0 \end{bmatrix}.$$

This is a simplification of equation (10) since, if, as our CPD model assumes, the hedonic slope coefficients are the same across areas for each specification characteristic so that $\beta_{iz}^a = \beta_{iz}^b = \beta_{iz}^0$, the coefficient on the difference between the share vectors of the two areas is a *characteristics-adjusted reference price vector* and no reference characteristics vector can be separately identified.

4.4 An Application to U.S. Labor Compensation Data

4.4.1 Data

The micro data used to construct the interarea indexes come from two sources: the employment cost index (ECI) and the Current Population Survey (CPS). The ECI data program produces quarterly indexes that measure changes over time in wages and salaries and in the cost of total compensation. These indexes are calculated from micro data collected for sampled jobs in sampled establishments. All jobs in nonfarm private industry and in state and local governments are within the scope of the survey, meaning that the occupation coverage of the survey is nearly complete. The micro data available include the mean hourly wage and mean hourly compensation costs for all incumbents in the sampled jobs. Other data elements describe job or establishment characteristics: the establishment's number of employees; whether the employment is full-time or part-time; and whether the job is covered by a collective-bargaining agreement. This study utilized the data for 18,486 sampled jobs for the fourth quarter of 1993 in nonagricultural private industry. Details of variable definitions, sample exclusion restrictions, and summary statistics for all data are in appendix A.

A shortcoming of the ECI is that it does not collect key variables that are widely believed to measure human capital: education and labor market experience. To obtain these variables, data from the CPS were merged with the ECI micro data. The CPS is a monthly survey of households that contains information about the demographic characteristics and employment outcomes of individuals. For current purposes, we used the three monthly surveys for the fourth quarter of 1993 and restricted our sample to employed individuals in nonagricultural private industry. We collected information on schooling, age, industry, occupation, and area of residence for a sample of almost 140,000 workers.

Merging the data from the CPS with the ECI presents a challenge because the ECI micro data contain the means for jobs while the CPS contains data for individuals and, of course, the individuals covered in the two surveys are not necessarily the same. The strategy we followed was to calculate weighted mean values for CPS variables for cells defined by local area, occupation, and industry. The industry and occupation cell classification used for this purpose was determined by the availability of data; we chose to create cells defined by local area, major occupation group, and six industry groups.

After matching the CPS cell-level data to the ECI micro data for individual jobs, we had to determine an appropriate locality/industry/occupation classification for the purposes of computing the interarea indexes. The methodology in the previous section, specifically equation (12), calls for estimating separate regressions for cells defined by industry and occupation in order to recover estimates of local area dummies for each cell. There is a trade-off between the size of the smallest local area for which we can calculate interarea indexes and how finely the industry/occupation cells can be disaggregated. We selected a set of cities that included both those that are the largest and those that are of interest in the federal pay-setting process. The remainder of the data were aggregated into census geographic divisions (as "rest of division"). We then determined that indexes could be calculated for these local areas using eighteen industry/occupation cells, defined by major occupation group and whether the job is in a goods- or service-producing industry.

To give the reader a feel for the underlying data, we present some summary data in tables 4.1 and 4.2. Table 4.1 gives wage and compensation shares and levels by our job classification scheme; major occupation groups are presented within the two broad industrial groupings. The first column, labeled *wage share*, reports the fraction of total wages that falls in the given category. These statistics are useful for showing where the bulk of the data reside. The second column simply reports the average hourly wage in the given cell (all figures are in nominal dollars). Roughly speaking, the professional, technical, and executive occupations have the highest hourly wages, production workers and operatives have average wages, and laborers and service workers have below-average wages. There is a noticeable difference between the broad industry aggregates, with average wages in any particular occupation group being higher in the goods-producing industries. The third column gives shares of total compensation. Goods-producing industries have higher shares of total compensation than of total wages, reflecting the fact that a higher fraction of compensation comes in the form of benefits for workers in those industries. This fact is apparent in comparing average wages, in the second column, to average hourly compensation costs, in the final column. Finally, one other obvious inference that can be drawn from this table is that, given the wide variation in wages and compensation costs across jobs, index numbers might be

Table 4.1 **Industry/Occupation Shares, Wages, and Compensation**

	Wage Share	Average Wage	Compensation Share	Average Compensation
Goods-producing industries:				
Professional/technical	.047	22.60	.047	31.98
Executive/administrative	.054	26.03	.054	36.84
Sales	.007	17.09	.007	22.92
Administrative support	.042	11.60	.044	17.08
Precision production	.094	15.80	.102	24.12
Machine operatives	.066	10.74	.075	17.04
Transport operatives	.029	12.76	.032	19.73
Laborers	.030	9.70	.033	14.78
Service workers	.008	14.08	.008	20.30
Service industries:				
Professional/technical	.145	19.33	.140	26.17
Executive/administrative	.101	20.95	.098	28.47
Sales	.095	10.12	.087	13.14
Administrative support	.120	9.93	.118	13.75
Precision production	.029	11.93	.028	16.28
Machine operatives	.009	8.18	.009	11.55
Transport operatives	.014	9.28	.014	13.18
Laborers	.027	7.03	.026	9.54
Service workers	.084	6.00	.079	7.89

Source: Winter 1993 employment cost index.

Note: Wage share and *compensation share* are wage and compensation shares in the nonhousehold, nonfederal, nonagricultural economy. *Average wage* and *average compensation* refer to average hourly wages and compensation in nominal dollars, where averages within industry/occupation class are weighted by ECI sample weights.

expected to yield very different results than simple interarea differences in average compensation rates whenever there are interarea differences in the distribution of jobs.

Table 4.2 gives employment shares and average compensation by local area. The compensation shares, showing each local area's compensation as a fraction of total U.S. compensation, give some idea as to which metropolitan statistical areas have relatively few ECI job quotes. Because of their small sizes, localities such as Charlotte and Columbus might be expected to have fairly noisy compensation index estimates. The "rest-of-division" localities, on the other hand, tend to be rather large. Comparing column 2 with column 1 shows that larger metropolitan areas tend to have the highest average compensation costs, the "rest-of-division" localities the lowest. The final column (*compensation relative*) gives average compensation in the local area relative to the overall average compensation level in the data. The range in these area relatives is quite large. At one extreme, compensation in the Detroit, New York, and San Francisco areas is approximately 134 percent of average compensation in the United States; at the other extreme lies the East South Central locality, with 73 percent

Table 4.2 **Compensation by Local Area**

Local Area and Rest of Regional Division	Compensation Share	Average Hourly Compensation ($)	Compensation Relative
		Northeast Region	
Boston	.029	19.75	116.4
Hartford	.009	20.60	121.4
New England	.016	14.04	82.8
New York	.116	22.73	134.0
Philadelphia	.027	19.93	117.5
Pittsburgh	.011	20.22	119.2
Middle Atlantic	.070	17.18	101.3
		North Central Region	
Chicago	.039	19.19	113.1
Detroit	.024	22.70	133.8
Cleveland	.010	17.33	102.2
Milwaukee	.008	16.34	96.3
Dayton	.007	17.46	102.9
Cincinnati	.006	15.89	93.7
Columbus	.006	15.99	94.3
Indianapolis	.005	18.59	109.6
East North Central	.062	14.23	83.8
Minneapolis	.011	17.35	102.2
Kansas City	.006	20.94	123.4
St. Louis	.006	18.27	107.7
West North Central	.052	14.74	86.9
		South Region	
Washington, D.C.	.026	21.84	128.7
Atlanta	.014	17.81	105.0
Miami	.009	14.45	85.2
Tampa	.008	13.89	81.9
Charlotte	.005	14.32	84.4
South Atlantic	.077	13.82	81.5
East South Central	.042	12.35	72.8
Houston	.020	19.21	113.2
Dallas	.018	17.17	101.2
West South Central	.056	14.00	82.5
		West Region	
Denver	.008	15.08	88.9
Phoenix	.007	16.24	95.7
Mountain	.028	13.26	78.2
Los Angeles	.060	20.02	118.0
San Francisco	.032	22.74	134.0
Seattle	.012	21.61	127.4
San Diego	.010	20.86	123.0
Portland	.008	18.66	110.0
Pacific	.039	15.42	90.9

Source: Winter 1993 employment cost index.

Note: Compensation share is the area share of compensation in the nonhousehold, nonfederal, nonagricultural economy. *Average compensation* is average hourly compensation in nominal dollars, weighted by ECI sample weights. *Compensation relative* is the ratio of the local area average compensation to the U.S. average compensation.

of overall average compensation. A comparison of these figures with the index numbers presented below will give some idea as to the importance of interarea differences in job characteristics for compensation cost comparisons.

4.4.2 Regressions

The first step in the construction of the interarea indexes was the estimation of log wage and log compensation cost regressions (eq. [12]). In addition to local area dummies, five sets of covariates were included as explanatory variables in the regressions to capture factors that affect worker productivity. Following a long tradition in the labor economics literature dating back to Mincer (1962) and Becker (1964), years of schooling, years of potential labor market experience (age minus education minus six), and potential experience squared were included. These measure the average amount of human capital possessed by incumbents in the job.

The labor literature has shown that wages are positively associated with establishment size. Brown and Medoff (1989) argue that part of this wage-size relation arises because large firms attract higher-quality workers (even after controlling for observable characteristics). In order to control for this in the present study, we include a set of eight establishment-size class dummies.

In the literature, unionization is claimed both to increase and to decrease worker productivity. The traditional view holds that unions lower productivity by imposing staffing requirements and other restrictive work practices that prevent firms from efficiently utilizing capital and labor (Lewis 1986; Rees 1989). A more recent literature argues that unions enhance worker productivity (Freeman and Medoff 1984). First, unions provide a collective voice that communicates workers' preferences. This lowers worker discontent and turnover, increasing firms' incentives to invest in job-specific human capital. Second, unions typically establish seniority rules that may promote an environment where more senior workers are willing to provide less senior workers with informal on-the-job training. Finally, unions may enhance worker morale, motivation, and effort. While the literature is ambiguous about the effect of unions on productivity, most studies show that unions increase wages. To capture the effects of unions on productivity in our regressions, we include a dummy indicating whether a job is covered by a collective-bargaining agreement.

The literature generally shows that, after accounting for observed differences in human capital, part-time workers earn less than full-time workers (see Lettau [1997] and the citations therein). For at least two reasons, this differential may arise because part-time workers are on average less productive than their full-time counterparts. First, it is argued that innately less productive workers are more likely to select part-time jobs. For example, more productive workers may find it advantageous to work more intensively if their wages reflect their productivity. Second, average productivity might be lower for part-time workers owing to fixed daily setup costs that are spread over more working hours for full-time workers. We include a part-time dummy in our regressions to capture these productivity effects.

It is important to note that, while the education and potential experience variables are widely viewed as measuring human capital, the other three variables—establishment size, unionization, and part-time status—may have effects on wages that are not strongly associated with labor productivity. All three represent or proxy to some extent characteristics of the labor services transaction in an industry and locality as much as the characteristics of the service itself. Although the nature of the transaction may have productivity effects, this is not a foregone conclusion. Large nonunion firms, for example, may pay higher wages simply to forestall unionization. Union wages may be higher simply because of union monopoly power. If the purpose of including explanatory variables in the regressions is to control for factors that affect productivity, then it is possible that our index factors overcontrol for some of these transaction and other effects. In the analysis that follows, we include all the explanatory variables in the regressions, but we provide a set of adjustment factors associated with the explanatory variables. These factors measure the contribution of each variable (or variable group) to the unadjusted interarea differential. One advantage of our methodology is that analysts can add back the differential associated with a variable if they judge that it is not appropriate to control for that variable. In tables 4.3 and 4.4 below, *adjusted* refers to interarea measures adjusted for differences in all our conditioning variables explaining wage and compensation variation. Future formats for interarea compensation data could reasonably include multiple summary columns of adjusted data corresponding to multiple subsets of conditioning factors to satisfy the interests of various users.

Eighteen regressions were estimated separately for wages and compensation costs. There is a regression for each of nine major occupation groups in either the goods- or the service-producing industries. The regressions for wages appear in appendix table 4B.1, while those for compensation costs appear in table 4B.2.

The adjusted R^2's for the regressions are comparable to or higher than those found for wage regressions estimated on individual micro data. The regressions typically explain between 20 and 50 percent of the variation of wages, while the corresponding range for compensation costs is 30–60 percent.

As expected by theory and found in most data, wages and compensation costs tend to rise with education. The returns to education are perhaps on average slightly smaller than would be obtained from person-level micro data. However, there are a few instances in our regressions where the education coefficient is anomalously negative (although not large relative to the standard error). This may arise in part owing to small sample sizes for some of the regressions. Further, it is important to stress that we have an imperfect measure of education that is measured as a cell mean from CPS data. Within a regression, education varies across areas and across some industry groups but does not vary for a given industry and area. It is likely that the education variable would perform better if it were collected for the same unit of observation as the wage and compensation data.

Previous empirical work has shown that wages display an increasing, concave profile with experience. This is often observed in our regressions as well, although there are a number of instances where the profile is convex and downward sloping at relevant experience levels. As with the education variable, problems with the experience variable might arise because its values are cell means whose source of variation for any regression is across areas and to a lesser extent across broad industry groups. The sample statistics indicate that the standard deviation of experience is much lower in our data than it is in micro data. This low variance is not unexpected but could indicate that the variable cannot discriminate well in explaining wage variation.

As expected, jobs that are covered by union contracts command higher wages than uncovered jobs, while part-time jobs tend to receive lower wages. The return to union contract coverage is on average higher in these regressions than estimates derived from older data (Lewis 1986). Finally, also as expected, larger establishments tend to pay higher wages, with especially notable premiums for establishments with five hundred or more employees. Comparisons of the establishment-size coefficients across industry/occupation groups are difficult because of substantial variability across those cells in the average compensation of workers in the omitted category (one to nine workers). However, the establishment-size coefficient point estimates typically rise with establishment size.

4.4.3 Interarea Indexes of Wages and Compensation for the United States

Table 4.3 gives our main results for interarea wage rate differentials. The second column of the table presents Törnqvist wage indexes that control for the composition of employment across nine major occupation groups and two industry groups but where there are no other adjustments for observed differences in worker or job characteristics. The index numbers are relative to the reference wage generated by our method (described in sec. 4.3 above) of making the bilateral comparisons transitive; one may loosely interpret 100.0 as average for the United States.[2] As an example, the first entry in the second column indicates that wages, adjusted for broad differences in employment but unadjusted for observed differences in worker and job characteristics, are 10.1

2. Actually, neither the area share-weighted arithmetic nor geometric average of these locality levels is generally equal to 100.0 because of the way the reference shares and prices are determined using the "minimum bilateral relative adjustment" criterion implicit in our regression approach, in concert with our observation weighting, which gives greater importance to records representing relatively large bilateral average expense shares. Bilateral ratios of the index numbers in tables 4.3 and 4.4 produce a transitive system of parities as provided by the objective of our algorithm but do not provide a particular-level normalization. Interpretation of these data as levels requires a normalization to, e.g., the national average level, much as a time series of price index numbers would be normalized to be 100.0 in a particular time period to align it with other data series so normalized for a given analytic purpose. The data in tables 4.3 and 4.4 are, therefore, valid for ranking localities in terms of labor services input price levels. As we note elsewhere in the paper, the weighted EKS/CCD method of determining the reference shares and prices of the transitive parities implicitly does normalize the regional geometric average level to 100.0.

Table 4.3 **Wage Indexes**

Local Area and Rest of Regional Division	Average Wage Relative	Wage Index	Education	Experience	Establishment Size	Full-Time/ Part-Time	Union	Adjusted Wage Index
				Northeast Region				
Boston	118.8	110.1	102.5	99.5	100.2	99.6	99.2	109.1
Hartford	120.9	111.9	99.6	102.6	102.1	96.0	101.1	110.6
New England	84.6	88.8	98.4	98.6	97.2	98.1	98.3	97.7
New York	132.7	128.0	100.6	100.7	100.0	101.1	100.6	124.2
Philadephia	116.2	111.2	101.4	99.7	99.5	100.0	101.0	109.4
Pittsburgh	118.1	108.8	102.4	103.6	101.7	100.9	102.2	97.8
Middle Atlantic	99.0	101.6	99.2	98.8	101.5	99.1	100.6	102.4
				North Central Region				
Chicago	111.9	108.8	101.3	100.3	99.1	100.6	101.7	105.6
Detroit	115.7	114.4	101.6	98.9	105.8	99.4	103.2	104.9
Cleveland	94.9	98.6	97.9	100.1	102.0	98.0	102.4	98.2
Milwaukee	96.8	101.2	97.7	99.5	100.3	100.9	100.8	102.0
Dayton	103.8	91.2	98.5	99.1	99.0	99.2	100.2	94.9
Cincinnati	95.2	100.9	100.3	99.4	100.3	100.5	99.4	101.2
Columbus	98.2	96.4	102.5	99.3	99.8	101.6	99.6	93.7
Indianapolis	106.8	101.0	98.1	106.9	98.9	102.9	101.1	93.5
East North Central	83.4	88.9	98.9	98.9	99.4	99.9	100.6	91.0
Minneapolis	100.6	107.0	102.2	97.9	100.0	99.5	102.1	105.1
Kansas City	120.7	110.9	103.6	99.2	99.0	102.3	101.6	104.9
St. Louis	101.3	95.1	104.5	99.8	98.4	101.6	101.2	90.2
West North Central	85.6	85.7	98.9	98.2	100.9	99.6	99.9	87.9

(continued)

Table 4.3 (continued)

Local Area and Rest of Regional Division	Average Wage Relative	Wage Index	Education	Experience	Establishment Size	Full-Time/ Part-Time	Union	Adjusted Wage Index
				South Region				
Washington, D.C.	127.5	112.9	100.9	100.7	102.0	100.9	100.4	107.5
Atlanta	104.2	105.1	100.7	101.2	101.6	102.2	100.6	98.7
Miami	89.2	92.4	97.7	100.5	98.8	100.0	98.5	96.6
Tampa	84.7	87.4	99.7	99.7	100.3	99.3	97.8	90.3
Charlotte	86.4	87.3	100.7	96.4	101.7	99.3	97.2	91.6
South Atlantic	83.4	86.6	98.6	99.3	99.3	99.9	97.8	91.2
East South Central	74.2	78.1	99.2	99.3	98.2	99.8	100.0	80.9
Houston	116.4	111.1	100.0	101.2	99.2	98.8	98.3	113.9
Dallas	102.4	97.0	98.2	99.4	102.1	102.4	100.1	95.0
West South Central	84.2	81.4	97.8	98.8	99.3	99.7	98.2	86.7
				West Region				
Denver	94.3	96.2	100.6	98.7	97.9	99.9	99.6	99.4
Phoenix	98.0	98.6	98.8	101.4	105.2	100.1	98.4	95.0
Mountain	80.4	87.6	100.7	100.5	98.8	98.8	98.2	90.4
Los Angeles	119.4	114.0	100.3	100.6	101.0	99.6	99.9	112.3
San Francisco	136.7	128.6	103.8	101.4	100.9	99.7	100.9	120.5
Seattle	127.9	118.2	102.4	101.5	102.0	102.1	102.3	106.8
San Diego	119.7	115.3	102.7	98.9	98.4	99.1	99.5	117.1
Portland	108.7	109.6	102.1	98.6	97.9	101.0	99.9	110.1
Pacific	91.7	101.5	100.2	98.9	99.2	100.1	100.8	102.2

Source: Winter 1993 employment cost index.

Note: The adjusted wage index in the last column is the wage index in col. 2 divided by the product of the characteristics indexes in cols. 3–7, normalized to base 100.

percent higher in Boston than in the United States as a whole. The amount of interarea variation in employment-adjusted wage indexes is striking, with numbers ranging from 128.6 for San Francisco to 78.1 for the East South Central rest-of-division locality. Generally, one tends to find that wages are higher than average in the larger CMSAs (consolidated metropolitan statistical areas) and along the West Coast; wage indexes are much smaller than average in the rest-of-division localities. Controlling for the composition of employment across industry and occupation by using a wage index tends to reduce interarea differentials as compared to the unadjusted wage relatives that appear in the first column of table 4.3. This is most clearly seen in figure 4.1, which plots the wage indexes against the average wage relatives. The figure contains a regression line through the data points (estimated with unweighted OLS) and a forty-five-degree line. If controlling for the composition of industry and occupation employment had no effect on interarea wages, then the regression line would have a forty-five-degree slope. Instead, the regression line is flatter than the forty-five-degree line, indicating that, in part, wages are low in low-wage areas because employment is more heavily concentrated in low-wage industries and occupations.

The rightmost column of table 4.3 gives adjusted wage differentials corresponding to the Törnqvist indexes calculated using the local area dummies ($\hat{L}_i^a$). Comparing the wage index column with the adjusted wage index column of table 4.3, the interarea variation in characteristics-adjusted wages is generally smaller than the interarea variation in wages alone. The standard deviation of the wage index is 12.2, as opposed to a standard deviation of 9.8 for the adjusted wage index. This can be seen graphically in figure 4.2, which plots one index against the other. The regression line through the plot is again less steep than the forty-five-degree line, indicating that controlling for worker and job attributes raises the wage index for low-wage areas and lowers it for high-wage areas. That is to say, some of the interarea variation about U.S. mean

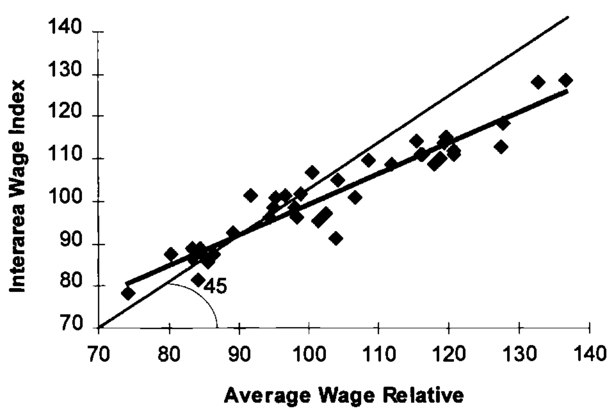

Fig. 4.1 Relation of interarea wage index to relative average wages

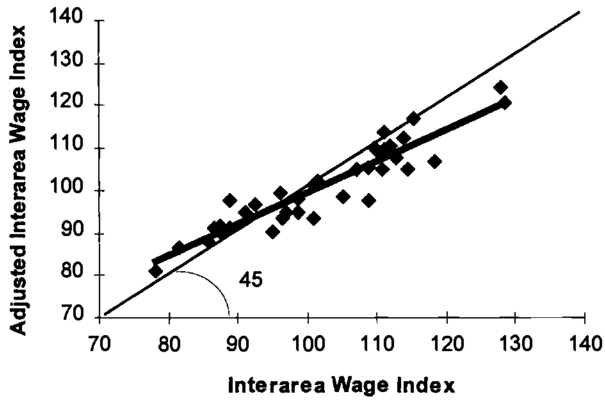

Fig. 4.2 Relation of characteristics-adjusted interarea wage index to interarea wage index

wages can be attributed to interarea differences in the observable characteristics *X*, even after accounting for interarea differences in industry/occupation employment distributions. One of the larger adjustments is in Detroit, where the observed differences in characteristics would imply an approximately 9 percent (114.4/104.9 = 1.09) premium to hire labor. The seven localities with the lowest employment-adjusted wage indexes (five of which are rest-of-division localities) have on average a 4.2 percent increase in wage indexes through our labor characteristics adjustments. In contrast, the seven localities with the highest employment-adjusted wage indexes have on average a 4.9 percent decrease through characteristics adjustments.

Whether the characteristics adjustment for a particular locality reflects a premium mainly attributable to larger establishment size, greater unionization rates, a more educated workforce, or some other reason is not clear from comparing the wage indexes. To address that question, table 4.3 contains a set of columns that report the contributions of the observable characteristics to unadjusted wage differentials. Recalling the discussion following equation (13), variations in the area relatives for a characteristic will be larger the larger is the coefficient for that characteristic in the wage regressions and the larger is the variation in the characteristic across areas. Our decomposition methodology implies that, for any given local area, the quality-adjusted index and the covariate contributions (appropriately scaled by one hundred) must multiply up to equal the employment-adjusted wage index. Whenever a number in one of the covariate columns exceeds 100 for a given area, the observed characteristic tends to raise wages in that area. For example, the covariate contribution for education in Boston, 102.5, indicates that Boston's workers are more highly educated than average, so their unadjusted pay is 2.5 percent higher than the average area owing to this characteristic. Although there is substantial noisiness in the results, especially among the smaller local areas, what one not surprisingly sees is that most of the adjustment factors for worker and job characteristics tend to exceed 100 for the highest-wage areas (e.g., San Francisco) and to fall short of 100 for the lowest-wage areas (including most of the "rest of regions"). Finally, as stated earlier, note that these adjustment factors can be used to add back to the characteristics-adjusted indexes the influence of variables that an analyst does not wish to remove in making interarea comparisons.

To get some sense of the relative contributions of the worker and job characteristics to wage differentials, figure 4.3 graphs the data from table 4.3 on area relatives for these characteristics against the interarea wage indexes. Each figure also plots the (unweighted) regression line through the points, and the scales of each figure are made the same to facilitate comparison. Each figure shows a positive correlation between the particular area relative and the wage index, indicating that, on average, all the characteristics contribute to interarea wage differentials. More significantly, the steepest regression line is for education, indicating that this characteristic is most important in explaining observed interarea differentials in the wage index. The union variable has the second steepest slope, while the part-time dummy variable has the flattest regression line. Since our wage regressions indicated that wages tend to be significantly lower for part-time jobs, the fact that this variable accounts for little variation in wages across areas stems from the fact that the proportion of part-time jobs varies little across areas.

Table 4.4 gives analogous calculations for hourly compensation, as opposed to wages. Given that wages constitute approximately 70 percent of compensation costs, it is not surprising to find that the gross patterns apparent in table 4.3 hold here as well. Controlling for industry/occupation and worker and job characteristics reduces interarea compensation differentials, implying that high-compensation areas receive high compensation partly for observable reasons. One difference between table 4.3 and table 4.4 is that the interarea differences in compensation indexes are slightly larger than those for wage indexes. One extreme example is Detroit, whose characteristics-adjusted compensation index (113.0) is much larger than its characteristics-adjusted wage index (104.9).

The greater interarea dispersion when computing compensation indexes holds for both the employment-adjusted and the characteristics-adjusted series. The compensation share–weighted standard deviations for these series are 17.2 and 13.3, respectively. Controlling for job and worker characteristics, therefore, reduces the interarea variation in compensation by about 23 percent. The greater interarea variation in compensation, as opposed to wages, no doubt reflects some combination of income effects (workers have income-elastic demands for health care, pensions, and other benefits) and tax effects (benefits are generally lightly taxed or not taxed at all, and the occupation composition of the labor force and income tax rates vary by locality). As employers making location decisions presumably care about compensation costs broadly defined, it is useful to know that interarea wage comparisons are likely to understate the interarea compensation differentials.

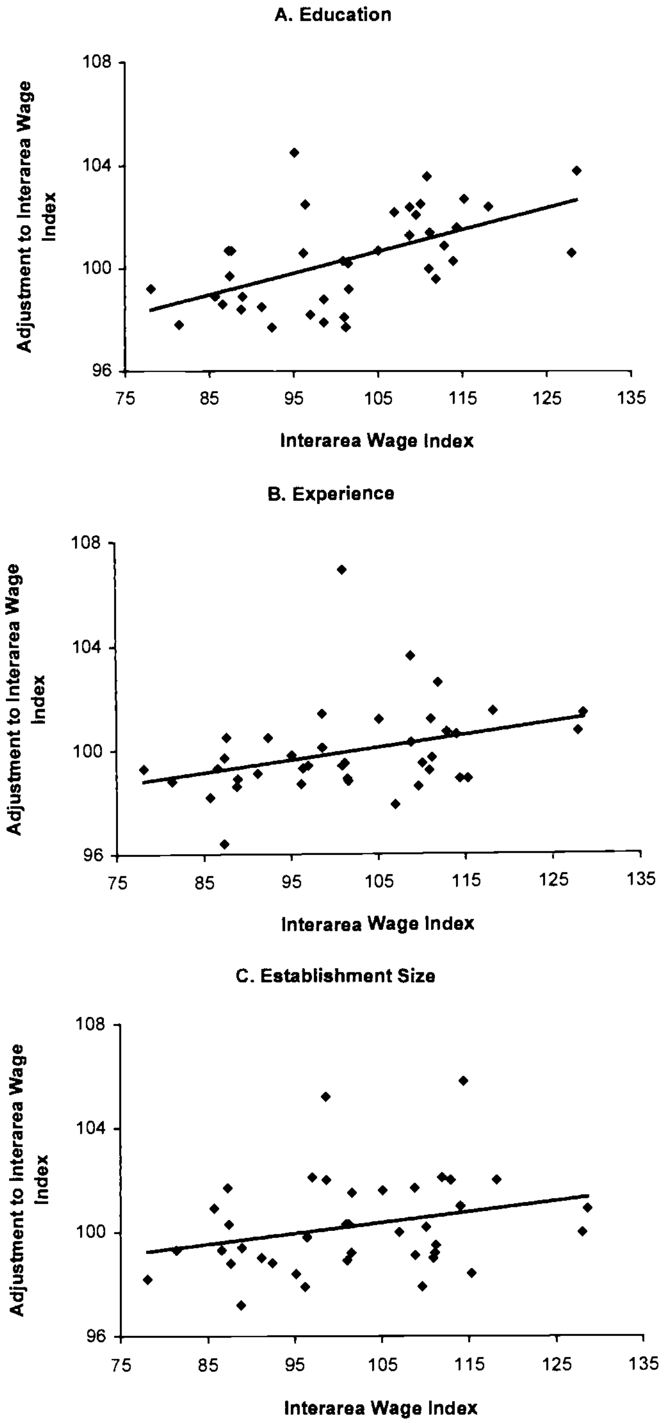

Fig. 4.3 Relation of labor services characteristics adjustment factors to interarea wage index: ***a*****, education;** ***b*****, experience;** ***c*****, establishment size;** ***d*****, full-time status;** ***e*****, union status**

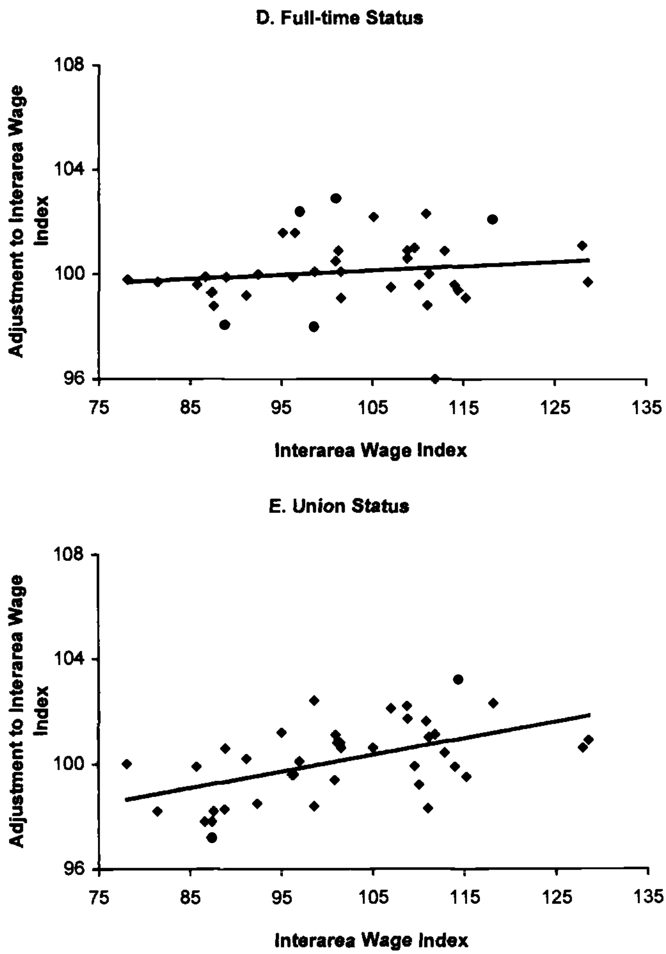

Fig. 4.3 (cont.)

4.5 Conclusion

We have applied a promising methodology for place-to-place price measurement to the problem of constructing interarea characteristics-adjusted compensation indexes, a methodology that blends hedonic regression and economic index number techniques. We have used a combination of establishment data from the BLS employment cost index program and household data on individual workers from the BLS Current Population Survey to provide a more complete picture of labor quality than has been available to analysts working with only household data. As would be expected, incorporation of the labor quality information generally reduced the variability of labor costs from place to place and provided insights into the contribution of various factors, such as educa-

Table 4.4 **Compensation Indexes**

Local Area and Rest of Regional Division	Average Compensation Relative	Compensation Index	Education	Experience	Establishment Size	Full-Time/ Part-Time	Union	Adjusted Compensation Index
				Northeast Region				
Boston	116.4	111.2	102.7	99.5	100.0	99.2	98.8	110.9
Hartford	121.4	115.5	99.9	102.7	102.4	95.1	101.6	113.9
New England	82.8	87.0	98.4	98.8	96.1	97.6	97.5	97.8
New York	134.0	130.3	100.7	100.9	99.8	101.4	101.1	125.2
Philadelphia	117.5	113.1	101.1	99.9	99.2	99.6	101.8	111.3
Pittsburgh	119.2	109.3	102.1	103.0	101.4	100.3	102.3	99.8
Middle Atlantic	101.3	103.2	99.2	99.0	102.1	99.0	100.9	103.0
				North Central Region				
Chicago	113.1	110.6	101.3	100.3	98.8	100.2	102.5	107.3
Detroit	133.8	129.5	101.3	99.1	109.1	99.6	105.1	113.0
Cleveland	102.2	104.4	98.5	100.5	103.1	98.3	103.7	100.3
Milwaukee	96.3	102.2	98.1	99.6	100.7	101.5	101.1	101.2
Dayton	102.9	92.3	98.0	98.9	99.1	99.5	100.2	96.3
Cincinnati	93.7	101.0	100.2	99.4	100.4	100.5	99.2	101.2
Columbus	94.3	93.6	102.0	98.7	100.1	102.1	99.5	91.5
Indianapolis	109.6	101.0	98.4	106.8	99.1	102.8	101.5	93.0
East North Central	83.8	88.9	98.7	98.8	99.4	99.8	100.9	91.0
Minneapolis	102.2	107.5	102.3	97.8	100.7	99.3	103.1	104.2
Kansas City	123.4	112.6	103.2	99.7	99.3	102.9	102.1	104.9
St. Louis	107.7	99.1	104.8	99.3	98.3	101.9	101.6	93.7
West North Central	86.9	85.8	98.8	98.1	101.0	99.6	99.7	88.3

	South Region							
Washington, D.C.	128.7	114.3	101.0	100.8	103.1	101.2	100.7	106.9
Atlanta	105.0	105.0	101.1	100.5	101.8	102.6	100.8	98.2
Miami	85.2	88.9	97.5	100.1	98.6	100.2	97.8	94.1
Tampa	81.9	85.5	99.9	99.9	100.1	99.6	96.8	88.8
Charlotte	84.4	84.7	100.7	96.7	101.6	99.8	96.0	89.4
South Atlantic	81.5	84.2	98.5	99.3	99.0	100.0	96.8	89.7
East South Central	72.8	76.4	99.2	99.3	97.9	99.8	99.8	79.6
Houston	113.2	109.3	100.0	101.6	99.3	98.2	97.4	113.1
Dallas	101.2	96.3	98.5	99.3	102.3	102.7	100.0	93.7
West South Central	82.5	79.1	97.7	98.7	98.9	99.8	97.4	85.3
	West Region							
Denver	88.9	91.2	101.3	98.4	97.5	99.3	99.5	95.0
Phoenix	95.7	97.1	98.8	100.7	107.1	100.2	97.6	93.2
Mountain	78.2	85.2	100.6	100.2	98.1	98.5	97.3	89.8
Los Angeles	118.0	112.8	100.3	100.5	101.3	99.6	99.8	111.2
San Francisco	134.0	127.8	103.7	101.4	101.1	99.6	101.6	118.9
Seattle	127.4	118.0	101.9	102.0	101.5	102.7	103.2	105.6
San Diego	123.0	119.9	102.2	99.1	98.8	99.0	99.5	121.7
Portland	110.0	111.6	101.8	98.3	97.0	101.0	100.1	113.7
Pacific	90.9	100.6	99.8	99.0	98.9	99.9	101.0	102.1

Source: Data from winter 1993 employment cost index; October, November, and December 1993 Current Population Survey.

Note: The adjusted compensation index in the last column is the compensation index in col. 2 divided by the product of the characteristics indexes in cols. 3–7, normalized to base 100.

tion, experience, establishment size, union status, and full-time work status, to the level of labor compensation in major urban centers of the United States.

Enhancements to the data are needed. Fortunately, there are prospective developments on this score. The BLS Office of Compensation and Working Conditions is currently undertaking a major redesign and integration of its three major compensation surveys, the employment cost index, the Employee Benefit Survey, and the Occupational Compensation Survey. One salutary result of this for the ECI is a substantial increase in sample size from the current five thousand establishments to at least twice that number. Of particular interest for interarea comparisons is the adoption of an area-and-industry-based rather than a solely industry-based rotational scheme for the samples in the new integrated survey, whose total size will be approximately thirty thousand establishments. Comprehensive data on job content is included in the list of data elements to be collected from all establishments in the survey, greatly expanding the number and explanatory power of the covariates that can be used for characteristics adjustment.

Appendix A
Data

The Employment Cost Index (ECI)

The ECI is a quarterly survey of randomly sampled establishments designed to produce estimates of wage and compensation cost changes. Within establishment, jobs are randomly sampled at the establishment initiation into the sample (sampling is carried out with probability proportional to establishment employment in the occupation). For each job, the ECI collects average wages and average compensation costs for the workers in the job. Nonwage compensation includes leave (sick leave, vacations, and holidays), supplemental pay (overtime, nonproduction bonuses), employer contributions to pensions and retirement savings accounts, health benefits, life and accident insurance, legally required labor expenses (state and federal unemployment insurance, workers' compensation, social security), and some other miscellaneous fringes. The ECI converts all data collected to a cost-per-hour-worked basis. The ECI micro data also attach various establishment or job characteristics to each job quote, including more detailed industry and occupation codes, establishment size, the job's work schedule, and whether the job is covered by a union contract. The ECI collects quarterly updates on the wages and compensation costs and uses these updates to compute quarter-over-quarter and year-over-year indexes of change. Establishments are replaced in the sample using an industry rotation; the entire sample is replaced over the course of four to five years.

For this study, we gathered a data extract from the ECI for the last quarter of 1993. We kept all private-sector job quotes for which we had valid wage and compensation data, meaning that the job quote was used in computing the ECI. Data can be invalid for two main reasons. The first is that the data represent the establishment's responses at initiation, which of course are not used in computing the most recent ECI change. We exclude these data mainly so our sampling weights remain approximately correct; including these observations would improperly overweight the industries that are the focus of initiation. The second is that establishments may be unable to, or may refuse to, report some benefits or wages for a particular job. In this case, the BLS attempts to impute wages or benefits on the basis of the nonmissing data available; cases where these attempts fail are essentially dropped from the ECI calculations. Finally, we note that, in some instances, the job's work schedule cannot be calculated and hourly compensation must be imputed even though the ECI has valid compensation data. Once exclusion restrictions are made, we have a sample of 18,468 job quotes.

Because sample replacement is made on an industry rotation pattern and sample weights are not adjusted through the life of the industry panel, normal sample attrition results in cross-sectional samples that overweight more recently initiated industries. Accordingly, we adjust the ECI sampling weights to bring them current by adjusting two-digit SIC employments to equal those published in the BLS *Employment and Earnings* series. References to the ECI sampling weights in the text and tables reflect this weighting adjustment.

The Current Population Survey (CPS)

While the ECI attempts to sample establishments randomly, the CPS is designed to sample addresses randomly and collect information on the households at each sampled address. The main function of these surveys is to generate official employment and unemployment statistics; however, they are utilized by researchers in a number of other ways as well. The survey is conducted monthly, with a given household surveyed for four months, not surveyed for eight months, and then surveyed for four final months, at which point the household leaves the sample. The survey collects demographics and current employment outcomes, among other items, for each person in a sampled household.

We pooled the October, November, and December 1993 CPS surveys to gather worker characteristics by industry, occupation, and local area at approximately the same time frame as our ECI data. The sampling design guarantees some overlap in the month-to-month samples, but that overlap does not imply redundant information in all cases because of changing employment rates, industry and occupation distributions, etc. Our sample exclusions were made primarily to maintain comparability with the ECI sample: we included only individuals employed by nonagricultural, private-sector employers. Our final sample contains 138,902 observations.

The covariates from the CPS data are mainly measures to proxy for human capital or other factors typically thought to affect wages. We have data on educational attainment, which we have converted into a measure of the years of schooling acquired by the individual. We derive "years of potential labor market experience," or approximate years out of school, as a proxy measure for the amount of general human capital acquired by the individual through work; it is defined as age − years of schooling − 6 (if less than zero, it is recoded to zero). Experience is entered as a quadratic to capture depreciation and decreasing investment rates through time (see Mincer 1974).

In order to match these data to our ECI sample, we averaged these covariates up to cell levels, where cells are defined by the area locations, six industry groups, and the nine major occupation groups. Averages are weighted averages, with weights being CPS sample weights. In matching the CPS to the ECI data, a small number of localities had missing values for some of the industry/occupation cells. These were allocated values from a donor cell of similar attributes within the local area. As these imputations account for a very small portion of the data, our results do not depend on the particular allocation method used.

Appendix table 4A.1 contains summary statistics, weighted by sample weights, for hourly wages, hourly compensation, and various job characteristics from the ECI.

Table 4A.1 Sample Statistics

Variable	Mean	SD	Minimum	Maximum
	ECI Variables			
Average hourly wage	12.07	9.31	2.13	277.38
Average hourly compensation	16.97	12.96	2.13	470.34
ln(hourly wage)	2.30	.59	.76	5.63
ln(hourly compensation)	2.62	.64	.76	6.15
Establishment size 10–19	.10	.29	0	1
Establishment size 20–49	.15	.36	0	1
Establishment size 50–99	.13	.34	0	1
Establishment size 100–249	.18	.38	0	1
Establishment size 250–499	.09	.28	0	1
Establishment size 500–999	.08	.26	0	1
Establishment size 1,000–2,499	.08	.28	0	1
Establishment size 2,500+	.09	.28	0	1
Works < 35 hours/week	.21	.41	0	1
Covered by union contract	.14	.35	0	1
	CPS Variables			
Years of schooling	12.95	1.25	4	17
Years of potential experience	17.39	3.87	0	48
Experience squared	466.54	149.31	0	2,304
Number of observations	18,468			

Source: Winter 1993 employment cost index; October, November, and December 1993 Current Population Survey.

Note: Data are weighted by employment cost index sampling weights.

Appendix B
Hedonic, Country-Product-Dummy Regressions

The regressions in equation (12),

$$\ln p_{ij}^{a} = X_{ij}^{a}\beta_i + L_i^{a} + \varepsilon_{ij}^{a}, \tag{12}$$

where p_{ij}^{a} represents the average hourly wage or compensation for the jth quote for job i in location a, X_{ij}^{a} represents data on the characteristics of the job and the worker, and L_i^{a} represents a local area effect for job i in area a, are essentially analogs to the country-product-dummy model in international product price comparisons. These regressions allow the coefficients on X_{ij}^{a} and the local area effects to vary across jobs. Tables 4B.1 and 4B.2 give weighted least squares estimates for (12), where the weights are the sample weights from the ECI.

It is worth discussing a few obvious points of interpretation. The coefficients give the estimated marginal effect on wages within the industry/occupation cell. One would expect these marginal effects to depend on how broadly or narrowly the cells are defined and to differ from the marginal effect from a regression over all industry/occupation cells. Furthermore, the CPS covariates are averages; a given CPS covariate value is not ECI quote specific as multiple ECI quotes have the same values attached. In that case, the proper interpretations to place on the marginal effects are less clear. For example, one tends to find higher wages in the ECI sample in locations and jobs where the population workforce (not the ECI sample workforce, as this aspect is unknown) is more highly educated. Does the schooling variable proxy for the ECI sample workforce's schooling, its cognitive abilities more generally, or some other factors that are also related to wages? This leads to another issue, namely, the question of which variables to use as regressors. Presumably the "proper" selection of covariates depends on what they proxy for as well as on the end purpose of the generated statistics. If the end purpose of the statistics is to inform business location decisions, then one would want to control for those factors that are productivity related or that capture labor cost premiums that do not reflect productivity differences but that can be avoided by prospective new firms. Although sensible readers might disagree with details of our specification, we feel that the covariates with the largest effects on the interarea wage indexes would fall primarily into these categories. Finally, the standard errors in tables 4B.1 and 4B.2 are likely to be biased downward for the CPS variables since the regression equation disturbances are correlated within groups (Moulton 1986, 1990). At this point, we are mainly interested in generating consistent estimates of the local area effects and are less interested in confidence intervals. Presumably, generating correct standard errors would be more straightforward if estimating the hedonic regressions and interarea indexes simultaneously were practicable.

Table 4B.1 Hedonic Wage Regressions

	Professional/ Technical	Executive/ Administrative	Sales	Administrative Support	Precision Production	Machine Operatives	Transport Operatives	Laborers	Service Workers
	A. Goods-Producing Industries								
Establishment size 10–19	1.02	−.03	.04	.12	.09	.25	−.24	.02	.41
	(.24)	(.11)	(.21)	(.05)	(.03)	(.07)	(.08)	(.07)	(.32)
Establishment size 20–49	1.13	0.24	.20	.06	.08	.08	−.14	.05	.65
	(.20)	(.09)	(.22)	(.05)	(.03)	(.07)	(.07)	(.06)	(.44)
Establishment size 50–99	1.11	.29	.61	.21	.17	.09	−.09	−.01	.48
	(.20)	(.09)	(.20)	(.05)	(.03)	(.07)	(.07)	(.07)	(.26)
Establishment size 100–249	1.14	.40	.29	.06	.15	.08	−.08	.02	.84
	(.20)	(.09)	(.19)	(.05)	(.03)	(.06)	(.07)	(.06)	(.27)
Establishment size 250–499	1.16	.28	.72	.20	.16	.08	.03	.07	.80
	(.20)	(.09)	(.22)	(.05)	(.04)	(.07)	(.08)	(.06)	(.28)
Establishment size 500–999	1.30	.43	.73	.19	.31	.21	−.16	.22	.85
	(.20)	(.09)	(.31)	(.05)	(.04)	(.07)	(.09)	(.07)	(.30)
Establishment size 1,000–2,499	1.32	.46	.47	.24	.30	.37	.09	.17	.98
	(.20)	(.09)	(.31)	(.05)	(.04)	(.07)	(.08)	(.07)	(.27)
Establishment size 2,500+	1.50	.47	.18	.30	.29	.50	.02	.27	1.10
	(.20)	(.08)	(.30)	(.05)	(.04)	(.07)	(.09)	(.08)	(.28)
Works < 35 hours/week	.14	−.40	−.80	−.27	−.12	−.19	−.08	−.17	.43
	(.08)	(.18)	(.17)	(.03)	(.06)	(.12)	(.07)	(.05)	(.11)
Covered by union contract	.23	.07	.16	.15	.20	.22	.23	.46	.31
	(.05)	(.10)	(.30)	(.02)	(.02)	(.02)	(.03)	(.03)	(.09)
Years of schooling	−.03	.02	.11	.02	.06	.00	.13	.06	.06
	(.04)	(.04)	(.07)	(.03)	(.02)	(.04)	(.03)	(.02)	(.05)
Years of potential labor market experience	.01	−.05	−.04	.03	−.01	.04	−.05	.01	.07
	(.03)	(.03)	(.04)	(.01)	(.02)	(.03)	(.02)	(.02)	(.02)
Experience squared/100	−.05	.08	.09	−.07	−.03	−.04	.13	−.06	−.13
	(.06)	(.06)	(.12)	(.03)	(.04)	(.07)	(.04)	(.05)	(.05)
Observations	829	675	97	1,128	1,649	1,208	339	500	124
Adjusted R^2	.32	.25	.58	.34	.37	.47	.43	.55	.75

	B. Service Industries								
Establishment size 10–19	−.24	.14	.19	−.05	.18	.01	.01	.24	−.10
	(.05)	(.05)	(.04)	(.02)	(.06)	(.18)	(.07)	(.05)	(.03)
Establishment size 20–49	−.14	.13	.18	−.04	.25	−.11	.18	.06	−.10
	(.05)	(.05)	(.04)	(.02)	(.05)	(.13)	(.06)	(.04)	(.03)
Establishment size 50–99	−.04	.21	.27	−.07	.38	.02	.22	.05	−.14
	(.04)	(.05)	(.04)	(.02)	(.06)	(.10)	(.07)	(.04)	(.03)
Establishment size 100–249	.01	.05	.09	−.04	.36	−.20	.24	.11	−.02
	(.04)	(.05)	(.03)	(.02)	(.07)	(.10)	(.08)	(.04)	(.03)
Establishment size 250–499	.04	.22	.05	−.02	.32	−.04	.12	.13	−.03
	(.05)	(.06)	(.05)	(.02)	(.10)	(.13)	(.09)	(.05)	(.03)
Establishment size 500–999	.09	.39	.08	−.03	.45	−.16	.35	.13	.03
	(.05)	(.06)	(.07)	(.02)	(.10)	(.13)	(.13)	(.07)	(.04)
Establishment size	.08	.32	.14	.01	.31	−.13	.22	.38	−.02
1,000–2,499	(.04)	(.06)	(.09)	(.02)	(.11)	(.23)	(.19)	(.06)	(.04)
Establishment size 2,500+	.05	.29	.04	.00	.42	.11	.20	.17	.15
	(.04)	(.06)	(.06)	(.03)	(.10)	(.20)	(.30)	(.08)	(.05)
Works < 35 hours/week	−.17	−.46	−.62	−.23	−.34	−.40	−.35	−.28	−.22
	(.02)	(.08)	(.02)	(.01)	(.08)	(.10)	(.06)	(.02)	(.02)
Covered by union contract	.06	−.26	.09	.12	.21	.25	.24	.29	.07
	(.04)	(.13)	(.05)	(.03)	(.07)	(.10)	(.06)	(.03)	(.03)
Years of schooling	−.01	.09	.29	.09	.01	−.07	.06	−.01	.01
	(.03)	(.03)	(.03)	(.02)	(.06)	(.06)	(.05)	(.02)	(.03)
Years of potential labor	−.01	−.01	.07	−.01	−.05	.02	.04	.01	.06
market experience	(.02)	(.02)	(.02)	(.01)	(.02)	(.03)	(.02)	(.01)	(.01)
Experience squared/100	−.01	.00	−.12	.03	.12	−.12	−.09	.00	−.11
	(.06)	(.05)	(.05)	(.02)	(.05)	(.07)	(.03)	(.03)	(.03)
Observations	2,058	1,482	1,692	3,259	425	165	246	659	1,933
Adjusted R^2	.21	.25	.48	.28	.32	.38	.45	.43	.41

Source: Winter 1993 employment cost index.

Table 4B.2 **Hedonic Compensation Regressions**

	Professional/ Technical	Executive/ Administrative	Sales	Administrative Support	Precision Production	Machine Operatives	Transport Operatives	Laborers	Service Workers
	A. Goods-Producing Industries								
Establishment size 10–19	1.11	−.01	.04	.17	.14	.30	−.15	.07	.32
	(.22)	(.11)	(.21)	(.06)	(.04)	(.08)	(.08)	(.07)	(.30)
Establishment size 20–49	1.21	.31	.28	.17	.14	.14	−.04	.16	.84
	(.19)	(.09)	(.21)	(.05)	(.03)	(.08)	(.07)	(.07)	(.42)
Establishment size 50–99	1.22	.33	.69	.30	.24	.13	.01	.03	.63
	(.19)	(.09)	(.20)	(.05)	(.03)	(.08)	(.07)	(.07)	(.25)
Establishment size 100–249	1.26	.42	.36	.18	.20	.15	−.02	.09	.93
	(.19)	(.09)	(.18)	(.05)	(.03)	(.07)	(.07)	(.07)	(.25)
Establishment size 250–499	1.28	.34	.82	.35	.23	.20	.14	.18	.93
	(.19)	(.09)	(.21)	(.05)	(.04)	(.07)	(.08)	(.07)	(.26)
Establishment size 500–999	1.48	.52	.81	.37	.40	.32	−.04	.30	1.01
	(.19)	(.09)	(.30)	(.06)	(.04)	(.08)	(.09)	(.07)	(.28)
Establishment size 1,000–2,499	1.48	.55	.53	.39	.41	.53	.22	.33	1.19
	(.19)	(.09)	(.30)	(.05)	(.04)	(.08)	(.08)	(.08)	(.26)
Establishment size 2,500+	1.66	.59	.20	.49	.43	.74	.33	.44	1.33
	(.18)	(.08)	(.30)	(.05)	(.04)	(.08)	(.09)	(.09)	(.27)
Works < 35 hours/week	.06	−.47	−.91	−.37	−.16	−.22	−.24	−.20	.31
	(.08)	(.18)	(.17)	(.04)	(.06)	(.14)	(.08)	(.05)	(.10)
Covered by union contract	.24	.19	.15	.21	.29	.29	.32	.56	.32
	(.05)	(.10)	(.29)	(.03)	(.02)	(.02)	(.04)	(.04)	(.09)
Years of schooling	−.04	.04	.11	−.01	.03	.00	.11	.06	.06
	(.04)	(.04)	(.07)	(.03)	(.02)	(.04)	(.03)	(.03)	(.05)
Years of potential labor market experience	.02	−.04	−.05	.04	.00	.03	−.05	.02	.06
	(.02)	(.03)	(.04)	(.01)	(.02)	(.03)	(.02)	(.02)	(.02)
Experience squared/100	−.07	.08	.14	−.08	−.02	−.01	.12	−.09	−.13
	(.06)	(.06)	(.12)	(.03)	(.04)	(.08)	(.04)	(.05)	(.05)
Observations	829	675	97	1,128	1,649	1,208	339	500	124
Adjusted R^2	.37	.32	.64	.41	.48	.56	.55	.62	.80

	B. Service Industries								
Establishment size 10–19	−.22	.15	.21	−.04	.23	.08	.04	.27	−.11
	(.05)	(.05)	(.04)	(.02)	(.06)	(.20)	(.07)	(.05)	(.03)
Establishment size 20–49	−.11	.15	.17	−.01	.30	−.11	.22	.09	−.08
	(.05)	(.05)	(.04)	(.02)	(.05)	(.14)	(.07)	(.04)	(.03)
Establishment size 50–99	−.01	.25	.28	−.05	.43	.10	.28	.09	−.09
	(.04)	(.05)	(.04)	(.02)	(.06)	(.11)	(.08)	(.05)	(.03)
Establishment size 100–249	.03	.07	.10	−.02	.37	−.13	.29	.16	.02
	(.04)	(.05)	(.03)	(.02)	(.07)	(.11)	(.08)	(.04)	(.03)
Establishment size 250–499	.07	.27	.06	.02	.41	.02	.17	.16	−.02
	(.05)	(.06)	(.05)	(.03)	(.10)	(.14)	(.09)	(.06)	(.03)
Establishment size 500–999	.16	.46	.13	.01	.52	−.06	.51	.17	.09
	(.05)	(.06)	(.07)	(.03)	(.10)	(.14)	(.14)	(.07)	(.04)
Establishment size	.16	.36	.22	.07	.41	−.18	.32	.42	.05
1,000–2,499	(.04)	(.06)	(.09)	(.03)	(.11)	(.25)	(.20)	(.07)	(.04)
Establishment size 2,500+	.10	.34	.09	.05	.46	0.25	.15	.18	.21
	(.04)	(.05)	(.06)	(.03)	(.10)	(.22)	(.31)	(.08)	(.05)
Works < 35 hours/week	−.26	−.53	−.70	−.35	−.42	−.55	−.47	−.34	−.27
	(.02)	(.08)	(.02)	(.02)	(.08)	(.11)	(.07)	(.03)	(.02)
Covered by union contract	.12	−.14	.19	.19	.30	.31	.37	.42	.16
	(.04)	(.13)	(.05)	(.03)	(.07)	(.11)	(.07)	(.04)	(.03)
Years of schooling	.00	.12	.29	.10	−.01	−.05	.03	−.02	.03
	(.03)	(.03)	(.03)	(.03)	(.06)	(.06)	(.05)	(.02)	(.03)
Years of potential labor	−.02	−.01	.07	.00	−.05	.02	.05	.01	.08
market experience	(.02)	(.02)	(.02)	(.01)	(.02)	(.03)	(.02)	(.01)	(.01)
Experience squared/100	.01	.00	−.12	.00	.09	−.13	−.10	−.01	−.14
	(.05)	(.04)	(.05)	(.02)	(.05)	(.08)	(.04)	(.03)	(.03)
Observations	2,058	1,482	1,692	3,259	425	165	246	659	1,933
Adjusted R^2	.26	.30	.53	.33	.38	.43	.56	.49	.50

Source: Winter 1993 employment cost index.

References

Becker, Gary S. 1964. *Human capital.* New York: Columbia University Press.

Brown, Charles, and James Medoff. 1989. The employer size–wage effect. *Journal of Political Economy* 97, no. 2:1027–59.

Caves, W., L. Christensen, and W. Diewert. 1982a. The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica* 50:1393–1414.

———. 1982b. Multilateral comparisons of output, input, and productivity using superlative index numbers. *Economic Journal* 92:73–86.

Diewert, W. 1976. Exact and superlative index numbers. *Journal of Econometrics* 4:115–45.

———. 1987. Index numbers. In *The new Palgrave dictionary of economics,* ed. J. Eatwell, M. Milgate, and P. Newman. New York: Stockton.

Fixler, D., and K. Zieschang. 1992a. Erratum. *Journal of Productivity Analysis* 2, no. 3:302.

———. 1992b. Incorporating ancillary product and process characteristics information into a superlative productivity index. *Journal of Productivity Analysis* 2, no. 2:245–67.

Freeman, Richard B., and James L. Medoff. 1984. *What do unions do?* New York: Basic.

Geary, R. G. 1958. A note on comparisons of exchange rates and purchasing power between countries. *Journal of the Royal Statistical Society* A121:97–99.

Khamis, S. H. 1970. Properties and conditions for the existence of a new type of index number. *Sankya* B32:81–98.

Kokoski, M., P. Cardiff, and B. Moulton. 1994. Interarea price indices for consumer goods and services: An hedonic approach using CPI data. Working Paper no. 256. Washington, D.C.: Bureau of Labor Statistics.

Lettau, Michael K. 1997. Compensation in part-time jobs versus full-time jobs: What if the job is the same? *Economics Letters* 56, no. 1:101–6.

Lewis, H. Gregg. 1986. *Union relative wage effects: A survey.* Chicago: University of Chicago Press.

Mincer, Jacob. 1962. On-the-job training: Cost, returns, and some implications. *Journal of Political Economy* 70, no. 5, pt. 2:50–79.

———. 1974. *Schooling, experience, and earnings.* New York: National Bureau of Economic Research.

Moulton, Brent R. 1986. Random group effects and the precision of regression estimates. *Journal of Econometrics* 32, no. 3:385–97.

———. 1990. An illustration of a pitfall in estimating the effects of aggregate variables in micro units. *Review of Economics and Statistics* 72, no. 2:334–38.

Rees, Albert. 1989. *The economics of trade unions.* Chicago: University of Chicago Press.

Summers, R. 1973. International comparisons with incomplete data. *Review of Income and Wealth,* ser. 19, no. 1 (March): 1–16.

Triplett, J. E. 1983. An essay on labor cost. In *The measurement of labor cost* (Studies in Income and Wealth, vol. 48), ed. J. E. Triplett. Chicago: University of Chicago Press.

Zieschang, K. 1985. Output price measurement when output characteristics are endogenous. Working Paper no. 150. Washington, D.C.: Bureau of Labor Statistics. Revised, 1987.

———. 1988. The characteristics approach to the problem of new and disappearing goods in price indexes. Working Paper no. 183. Washington, D.C.: Bureau of Labor Statistics.

Comment Joel Popkin

Summary of Paper

Broadly speaking, Pierce, Ruser, and Zieschang construct a spatial version of the employment cost index (ECI) for wages and for compensation with two differences in method from the time-series ECI. One is their use of transitive (or multilateral) Törnqvist place-to-place index number techniques. The ECI, by contrast, is neither superlative nor transitive. The Törnqvist method used by Pierce, Ruser, and Zieschang is derived from previous work by Caves, Christensen, and Diewert and by Kokoski, Moulton, and Zieschang. The second difference in method is that Pierce, Ruser, and Zieschang adjust the Törnqvist by using a hedonic regression incorporating both ECI and household survey data to control for labor composition effects. The ECI maintains looser controls in this regard by estimating separate indexes by industry, occupation, bargaining status of workers, etc. The two improvements have the effect of narrowing the spread in wages and compensation across the United States. For wages, just using the Törnqvist accounts for a major portion of the narrowing. For compensation, each of the two improvements narrows most measures of the spread by about equal amounts. For one measure, however, the regression accounts for more of the narrowing.

In their paper, Pierce, Ruser, and Zieschang present the two spatial indexes of the ECI that they have constructed, that is, the unadjusted and the regression-adjusted Törnqvist index. The unadjusted Törnqvist index that they present controls for nine major occupations and two major industry groupings (goods and services) in thirty-nine areas of the United States. In Pierce, Ruser, and Zieschang's parlance, that index is a composition-unadjusted index. The adjusted Törnqvist is adjusted for labor composition effects. For purposes of comparison, they also present a third index, which is a simple spatial index of wages and compensation relatives based on ECI sample weights for the same thirty-nine areas.

The method used for constructing the Törnqvist index adjusted for labor composition effects is based on micro data on jobs from the ECI and on the household survey from the Current Population Survey (CPS). ECI data are combined with CPS data to estimate hedonic regressions similar in specification to the country-product-dummy approach. The regression coefficients, that is, estimates of labor characteristic prices and local area effects, form the inputs to a multilateral Törnqvist index number construct. The unit of observation for the regression is a job. This is the unit priced in the establishment-based survey that underlies the ECI. The labor composition variables in the regression include ECI data on establishment size and the union and part-time status of the job priced.

Joel Popkin is president of Joel Popkin and Company, an economic consulting company, and was director of the former Washington office of the National Bureau of Economic Research.

Table 4C.1 A Comparison of the Three Interarea Indexes in the Pierce, Ruser, and Zieschang Paper

	Wages			Compensation		
	R	T	T^A	R	T	T^A
High values	136.7	128.6	124.2	134.0	130.3	125.2
Low values	74.2	78.1	80.9	72.8	76.4	79.6
Spread	62.5	50.5	43.3	61.2	53.9	45.6
Ratio to R	...	.81	.69	...	.88	.75
High/low	1.84	1.65	1.54	1.84	1.71	1.57
SD	19.49	15.61	12.58	20.44	17.18	13.28

The list of independent variables is enhanced by the use of CPS data to add the following additional variables to the regression: schooling, experience, part-time status, and union coverage status. Since the unit of observation in the CPS is the individual worker, the CPS variables in the hedonic regression represent group means computed over workers falling within specified area, occupation, and industry cells. These group means are mapped onto the corresponding observations on jobs from the ECI. It appears that 2,106 cell means from the CPS were mapped to 18,486 ECI job-price quotes. Thus, the mapping is not one to one; each CPS cell mean was, on average, matched to nine ECI job-price quotes.

Table 4C.1 shows the effect of using the unadjusted and adjusted Törnqvists. The table compares the three indexes in Pierce, Ruser, and Zieschang's paper. The column labeled R is the simple spatial relative, the column labeled T is the unadjusted Törnqvist, while the column labeled T^A is the adjusted Törnqvist. The high and low values presented in the table for each of the indexes are the highest and lowest values of the indexes. The highest values, however, do not always represent the same area. For example, the highest value for wages for R and T is for San Francisco, while the highest value for T^A is for New York.

From the estimates in Pierce, Ruser, and Zieschang's paper and supplemental data supplied by them, there are four ways of looking at how T and T^A improve spatial comparisons relative to R. One is by looking at the difference between the high and the low values of each index. That difference is labeled *spread* in the table. Another way is by looking at the ratio of the spread of T and of T^A, respectively, to the spread of R. It is labeled *ratio* in the table. A third way of looking at the improvement is from the ratio of the high to the low values of each of the indexes; it is labeled *high/low*. The final and most meaningful way of looking at the improvements is from the standard deviation of each of the indexes. That is labeled *SD* in the table.

From table 4C.1, it is apparent that all three of the indexes that the authors construct separately for wages and compensation show a wide dispersion across the United States. What the table demonstrates, however, is that the dispersion is narrowed by the use of T^A in comparison to T and R. One conclusion that can be drawn from the smaller dispersion of T^A relative to T and R is that labor composition has a sizable influence on estimates of wage costs and compensation cost relatives across areas in the United States.

For wages, a comparison of R to T shows that just moving from the simple spatial ECI to the unadjusted Törnqvist reduces the spread of high and low values by twelve index points (from 62.5 to 50.5). The adjustment to the Törnqvist narrows the spread by an additional 7.2 index points (from 50.5 to 43.3). Overall, when moving from the simple spatial ECI to the adjusted Törnqvist, the spread in wage costs is reduced by somewhat less than one-third ($1 - 0.69 = 0.31$ [from the line labeled *ratio*]). Another way of looking at the narrowing of the spread is to observe the relative costs of the highest wage area to the lowest one (the *high/low* line in the table). Before any adjustment for composition effects, the highest-cost area is 84 percent more expensive than the lowest-cost area. After application of the Törnqvist and the adjustment to it, the highest-cost area is only 54 percent more expensive in terms of wages, with most of the narrowing accounted for by moving to the Törnqvist alone. A comparable effect is observed for the standard deviation. Moving from the simple spatial ECI to the Törnqvist reduces the standard deviation by about 20 percent, while moving to the adjusted Törnqvist reduces it by an additional 15 percentage points.

To sum up the effects of the adjustments made by Pierce, Ruser, and Zieschang in measuring wages spatially, the use of T in place of R narrows considerably the dispersion of wages in the United States; the use of T^A in place of T narrows the dispersion further but by somewhat less.

For compensation, a similar narrowing is observed in the dispersion. Specifically, the measure labeled *spread* in the table shows that moving from the spatial ECI to the Törnqvist narrows the spread by 7.3 index points (from 61.2 to 53.9), and that is comparable to the narrowing of 8.3 points (from 53.9 to 45.6) when moving to the adjusted Törnqvist. The measures labeled *ratio* and *high/low* also show about equal narrowings between R and T and between T and T^A. For the standard deviation, the result is somewhat different. The standard deviation shows a greater narrowing between T and T^A than between R and T (19 vs. 16 percent). For compensation, then, the improvements made by the authors to the spatial indexes are by some measures split about evenly between just using the Törnqvist and using the adjusted Törnqvist. When, however, the standard deviation is used to assess the improvements made by the authors, the adjusted Törnqvist shows a larger improvement over the unadjusted Törnqvist than the latter does over the simple spatial measure.

The differences observed between wages and compensation in the extent to which the unadjusted Törnqvist and the adjusted Törnqvist account for the narrowing of the spatial indexes seem to suggest the following: that labor composition effects account for more of the difference in benefits than they do of differences in wages.

A more micro analysis of the labor composition effects shows that the key variables that Pierce, Ruser, and Zieschang used to explain wage cost differences across areas are establishment size, union coverage, schooling, and experience. However, a particular labor composition effect found to be important in one area is not found to be important in other areas. For example, establishment size and union status effects are key factors explaining the results for Detroit but not for New York, where full- versus part-time status appears more important.

For wages, an interesting result in the paper is that, the higher the level of R in an area, the higher is T^A *in percentage terms* for that area. For example, New York's R is 132.7, while the ratio of T^A/R in New York is 1.068 (a 6.8 percent adjustment). By contrast, in East South Central, R is 74.2, while its ratio of T^A/R is 0.917 (a -8.3 percent adjustment). This result points to an interesting asymmetry in the data: above-average wage costs are symptomatic of superior labor quality, while below-average wage costs are symptomatic of factors not related to labor quality. This asymmetry is worth further investigation. Is it, for example, an artifact of the methodology? Would this finding be affected if the regression methodology allowed labor characteristic prices to vary across areas? Is the finding consistent with what is known about local area labor market demand and supply conditions?

Comment 1: Work versus Worker

The paper needs to be mindful of the distinction between work (or the job) and the worker in the job. Is it the intent of the paper to estimate how compensation varies across areas in the United States for work having similar characteristics? Or does the paper intend to measure relative compensation across areas for workers having similar characteristics? In the former case, the hedonic regression should include only characteristics that are job descriptive, whereas, in the latter case, the regression should be limited to characteristics that are worker descriptive.

If the intent is to explain *all* interarea wage dispersion and then decompose it into job characteristic and work characteristic effects, it becomes even more important to be mindful of the distinction between the two. One distinction between the two is that worker characteristics are portable; that is, they move with the worker. Some variables appear to be both worker and job descriptive, but it is important to note that their interpretation changes depending on their use. For example, union coverage should be used as a job descriptive variable only if one believes that workers sort themselves randomly into union and nonunion jobs. But, if there are systematic differences in skills required for union and nonunion jobs, then union coverage should be used as a worker characteristic.

The existing literature should be referenced more fully to sort through these choices. For example, the seminal work by Brown and Medoff (1989) considered both worker- and job-descriptive aspects of employer size and their effect on wages. Brown and Medoff found support for the view that the employer size–wage effect is a consequence of differences in worker quality but little support for the view that the wage effect is a consequence of differences in job-related characteristics across employers of different size.

There is also a growing literature within labor economics that attempts to sort through the relative effects of job and worker characteristics on wages. Unfortunately, the variables typically used as job descriptive[1] in this literature are not currently available in either the ECI or the CPS. However, this literature can be fruitfully consulted if Pierce, Ruser, and Zieschang wish to decompose interarea wage dispersion into work and worker components. Their present method of distinguishing between ECI and CPS covariates is inadequate in this regard.

Comment 2: The ECI versus the CPS

The authors have given primacy to the ECI data, retaining the job as a unit of observation and molding the CPS data to fit this requirement. The reasons for this appear to be a desire to produce a spatial analog to the current ECI and the availability of data on benefits in the ECI. However, this approach is not as inexpensive as it is presented to be.

The authors extract three explanatory variables from the ECI—establishment size, union coverage, and part-time status. As discussed above, none of these variables is uniquely interpretable as job descriptive. In fact, there is considerable evidence pointing to their status as worker-descriptive variables. If data are available only on characteristics of workers, the appropriate choice of a unit of observation is a worker, not a job. And the proper data to use are household survey data, not establishment data.

The ECI variables used by Pierce, Ruser, and Zieschang are all recorded in the CPS on an annual basis. The ECI data could be used as a supplemental source of information. Data on nonwage compensation could be aggregated within cells defined by detailed occupation, establishment size, union status, region, etc. and merged with the CPS. Also, the merger of the two data sets in this manner is likely to be less costly from the point of view of the regression analysis than the gross aggregation of the CPS data presently adopted by the authors.

The main problem with the CPS data would be that establishment size is available only on an annual basis. Thus, quarterly indexes could not be computed. There may also be a loss in timeliness. Nonetheless, the advantages of the CPS data are sufficient to argue that, at least at this stage of the development of the index, the authors should also estimate a CPS-based interarea wage-cost index.

1. These are mostly *Dictionary of Occupational Titles*–type variables attempting to measure job objectives and complexity.

Comment 3: The Ultimate Uses of the Index

The paper alludes to two possible uses of the index. First, it suggests that the index could be used to inform employers' plant location decisions. From that point of view, the index would probably be more useful if disaggregated by industry and employer size rather than by industry and occupation. Second, the paper suggests that the index could be used to compare wage levels to implement the Federal Employee Pay Comparability Act (FEPCA). From this point of view, the index would be better served by relying more on CPS data and using more worker-descriptive variables. A worker-descriptive variable that immediately comes to mind in the context of FEPCA is gender.

Whether gender should be used as a variable in a hedonic wage regression has been previously debated in the literature. It is sufficient to note here that some of the most significant contributions to the field of measuring labor composition/quality have opted to use gender as an explanatory variable. These include the seminal work by Gollop and Jorgenson (1983) and U.S. Department of Labor (1993), the BLS's own index of labor composition (produced by its Office of Productivity and Technology). Note that the use of a gender variable would probably rule out the use of a job as the unit of observation.

Comment 4: Regression Results

The merger of the ECI and CPS data produces a variety of oddities in the regression results. One oddity is that the CPS variables are cell means; hence, the coefficients attached to them cannot be interpreted as characteristics prices (on the margin). Do these coefficients belong in the hedonic index? Another oddity is that, among the eighteen regressions estimated, about one-third have the wrong sign on the schooling variable. Yet another oddity can be seen in the coefficient of the experience variable. It appears to be significant in only seven regressions and has the wrong sign in two of these cases. A final oddity is in two of the ECI covariates. Specifically, wages appear to decline with establishment size in some regressions. These anomalies and the interpretation of CPS-variable coefficients as characteristic prices are issues that need to be analyzed in greater detail by the authors.

References

Brown, Charles, and James Medoff. 1989. The employer size–wage effect. *Journal of Political Economy* 97, no. 2:1027–59.

Gollop, F., and D. Jorgenson. 1983. Sectoral measures of labor cost for the United States, 1948–78. In *The measurement of labor cost* (Studies in Income and Wealth, vol. 48), ed. J. E. Triplett. Chicago: University of Chicago Press.

U.S. Department of Labor. Bureau of Labor Statistics. 1993. *Labor composition and U.S. productivity growth.* Bulletin 2429. Washington, D.C.: U.S. Government Printing Office.