CBA·NAU
College of Business Administration Northern Arizona University
Box 15066
Flagstaff AZ 86011

An Analysis of Average Annual Pay and Cost-of-Living in U.S. Metropolitan Regions

Working Paper Series 02-28— September 2002

Ronald J. Gunderson
Professor of Economics

James V. Pinto
Professor of Economics

College of Business Administration
Northern Arizona University
Box 15066
Flagstaff, AZ 86011

CBA · NAU
College of Business Administration
Northern Arizona University
Box 15066
Flagstaff AZ 86011

An Analysis of Average Annual Pay and Cost-of-Living in U.S. Metropolitan Regions

Ronald J. Gunderson and James V. Pinto

INTRODUCTION

Economists and regional scientists have produced extensive research to address issues of pay differentials across regions in the U.S. and in other nations. The research often centers on whether or not wage or salary differentials across regions converge, diverge or remain constant over time (Montgomery 1992; Eberts and Schweitzer 1994). Neoclassical models predict a narrowing of this differential over time providing markets are competitive and resources are mobile; yet empirical results suggests this has not occurred.
Furthermore, the presence or absence of selected amenities has been incorporated into studies in order to predict wage levels (Gabriel and Rosenthal 1999). Older work by Tiebout (1956) and more recent work by Epple and Romer (1991) have found that the migration decision is based in part on location-specific amenities implying that wage rates will fall when the value of local amenities increases in the absence of barriers to mobility. Other studies have investigated whether wage differentials across labor markets are influenced by relative differences in the cost of living (Dumond, et al. 1999).

Conclusions have varied depending on the time of the study, the locations chosen and selection of the predictor variables. Future studies will likely continue to generate similarly conflicting results; however, the value added of each study lies with the ongoing refinement and adjustments that compare and contrast the conditions under study in each analysis.

This paper examines the relationship between "average annual pay" in 2000 for 219 urban areas and the cost of living in these same locations. Regression analyses were undertaken for urban areas where one or more cities in an area participated in the ACCRA Cost of Living Index in 2000. The majority of the observation units included in this study are congruent with the MSA's (Metropolitan Statistical Areas) in the U.S. However, there are numerous exceptions since the ACCRA Index is not constructed along the lines of MSA boundaries. As examples, Minneapolis and St. Paul are presented as separate units of analysis for this study even though they each reside in the same MSA. Raleigh, Durham and Chapel Hill comprise one MSA; however, each market was analyzed as a separate entity for this paper,.

The first equation performs an analysis for the entire 219 communities included in this study. The populations in these locations accounted for 66 percent of the total U.S. population in 2000. Additional regressions were undertaken to analyze urban areas contained in each of the four separate U.S. Census Regions. The equations provide us with information about the strength of the relationship between annual pay and cost-of-living for different regions within the nation, and to test the hypothesis that similar relationships exist for all the Census Regions across the nation. Other variables were included in the equation to control for population size, climate and education in order to better define this relationship between annual pay and cost of living,

DEFINITION OF THE VARIABLES

One of the key elements in the decision to relocate or migrate from one location to another is whether or not the migrant's economic status is augmented by the move. The potential for enhancement in economic status is assumed to play a significant role in the decision as to whether or not to migrate. Therefore, it is important to know if relocating to a new area improves the economic status of the individual migrant.

As referenced earlier in this article, several researchers have attempted to answer this question by comparing the wages paid in each region with the relative cost-of-living in the same area. A person armed with this information could then determine, for example, if a 20 percent increase in wages commensurate with a move to a new location purchases 20 percent more goods and services, or if the 20 percent wage hike results in the purchase of greater or lesser percentage amounts of these items.

2000 Annual Pay

Our analysis does not focus on the wages paid in a region. Instead, we use the Average Annual Pay for 2000 for all Covered Workers by Metropolitan Area. The Bureau of Labor Statistics (BLS) at the U.S. Department of Labor compiles and publishes this data series. In 2000, the annual pay in metropolitan areas averaged $36,986.[1]

The "average annual pay" is the dependent variable in the regression equations. Employers submit information on annual pay to the BLS by taking the total annual payroll for employees covered by unemployment insurance and dividing it by the average number of monthly employees. An average pay figure is reported for each of the metropolitan areas in the U.S. Since we often included more than one city from any single metro area in our study, we assigned the value of the average annual pay that existed in the entire metropolitan area to each city.

ACCRA Index

The first of the independent variables is the ACCRA Index. This index, prepared by the American Chamber of Commerce Researchers Association, relies on a common methodology across cities and towns to provide information concerning the cost of purchasing selected items in various metropolitan and nonmetropolitan regions. This cost of living index is our key variable in supporting our hypothesis that a significant relationship exists between annual pay and cost of living differentials for each locality. The Index includes quarterly price information on housing, grocery items, utilities, health care, transportation and miscellaneous goods and services, as well as a composite index. The overall U.S. mean value for each component of the Index is set at 100.0. For this analysis, we have used an unweighted average of the quarterly indexes for 2000 for the composite index. The ACCRA index is not available for all metropolitan statistical areas (MSA's), and even then, it is only available for participating jurisdictions within an MSA, not the entire MSA. However, data that address cost of living or price differentials at this level of geography are typically unavailable (Glaeser 1998). Since the ACCRA index provides the key indicator to express the relationship between annual pay and cost of living, we limited the number of observations in this study to those areas (cities, counties, MSA's) where we could obtain both the annual pay and the ACCRA observation for 2000.

Area Population

Several different proxies were considered for use as an area population variable. The best results were obtained when a dummy variable was employed that distinguishes between small urban areas and large urban areas.

If population in the area $< 500{,}000 = 0$

If population in the area $\geq 500{,}000 = 1$

The selection of 500,000 residents as the separator between "small" and "large" metropolitan regions was due to various considerations. The decision to use this figure represents our feeling that cities with populations below this threshold may to some extent, still generate the sense of a smaller community in contrast to the overwhelming presence of a major metropolitan entity that eventually begins to emerge as the area population continues to grow. Although the decision to use 500,000 as the separator was arbitrary, a review of the qualities of some of the communities with populations closely above and below this number was partially responsible for the use of this breakpoint.

Climate Index

The climate index is the one adopted by *The Places Rated Almanac.* This almanac ranks each of the metropolitan areas on a number of variables with the goal of identifying the "best places to live in the U.S. and Canada." The climate variable represents a compilation of information derived from four weather-related factors – winter mildness, summer mildness, seasonal affect and hazardousness. The factors are compiled into an index where each metro region's score represents its percentile on a scale ranging from a low of 0 to a high of 100. A region that scores high on the climate index experiences minimal temperature extremes in the winter and summer seasons and has insignificant amounts of windchill and humidity. The seasonal affect is a proxy for the negative psychological impacts associated with excessive cloud cover and precipitation. Hazardousness is a measure of the negative and potential impact of injury and death that can occur due to snowfall, thunderstorms and strong winds.

---

[1] Average Annual Pay in Metropolitan Areas, 2000. Bureau of Labor Statistics. U.S. Department of Labor. Washington, D.C. http://www.bls.gov/cew/

EDUCATION INDEX

The education index is also the one adopted by *The Places Rated Almanac.* A number of variables are included in this index including school support, library popularity, college town and college options. School support includes information related to both public and private schools on the number of pupils per classroom and the percent of funding that originates from local rather than state or federal sources. Fewer students per classroom and more local funding generate a higher score. Library popularity is a ratio calculated by adding the number of volumes contained in local libraries to the usage or circulation of those volumes. This figure is then divided by the metro area population to obtain the library popularity index. The remaining components of the index reflect the presence or absence of a college or university in the region as well as the size and types of programs available and the opportunity for residents to meet their educational needs at a variety of higher education institutions including night and weekend instruction, graduate programs and occupational certificates. The education index was constructed similar to the climate index, and each metro region's score represents its percentile on a scale ranging from a low of 0 to a high of 100.

EMPIRICAL RESULTS

A best-subsets model using the PHStat 1.4© add-in to Excel 2000© was used to narrow a field of potential independent variables to those presented in all tables. The best-subsets approach evaluates the best subsets of models for a given number of independent variables or all possible regression models for a given set of independent variables. Other independent variables considered for this analysis included the actual population of the areas, the actual population squared and the population growth in each area from 1990 to 2000. SAS 8.0 for Windows© was used to confirm the Excel analysis, to determine the variance inflation factors (VIF) and to complete White's Test (White 1980b) and the Breusch-Pagan test (Breusch and Pagan 1979).

For all the models in Table 1 through Table 5 the overall F tests for the independent variables taken as a whole were highly significant with observed levels of significance (p-values) less than 0.01 in each case.

The VIFs in all the models revealed values less than 10 which indicate the absence of significant multicollinearity among the independent variables. If a set of explanatory variables is uncorrelated, then the individual VIFs will be equal to one. Most of the VIFs in the following models have values between 1 and 2. White's Test and the Breusch-Pagan test are used to test for heteroscedasticity. These tests are particularly important when using cross-section data in the models. All observations on the data were taken in 2000. White's test is general because it makes no assumptions about the form of the heteroscedasticity. The Breusch-Pagan test assumes that the error variance varies functionally with a set of regressors. In general, the observed level of significance for the White's is larger than that for the Breusch-Pagan test. In both of these tests, the null hypothesis is homoscedasticity; thus, we wish to fail to reject the null hypothesis with large p-values. We take the position that if the observed level of significance is greater than 0.10 for either or both tests, then there is no significant heteroscedasticity in the model. This result is found in all the models presented below.

ALL AREAS

Table 1 shows the results of analysis in all 219 areas considered in the study. Each of the independent variables is significant at varying observed levels of significance (p-values). All of the coefficients on the independent variables have the *a priori* expected signs. We expect the ACCRA Index, Area Population and Education index to all have positive signs. We postulate the sign on the Climate Index to be negative. The coefficient of multiple determination is 0.623 for this model.

Table 1: Full Model: All Areas Regression

R Sq = 0.623 Adj. R Sq = 0.616 n = 219

White's Test 0.2135 Breusch-Pagan 0.0754

Predictor	Coefficients	t Stat	P-value	VIF
Intercept	7285.19	3.995	8.87883E-05	NA
ACCRA Index	194.13	10.394	9.4941E-21	1.149
Area Population	2258.20	3.632	0.000352193	1.838
Climate Index	-16.21	-1.844	0.066513339	1.1805
Education Index	66.069	6.241	2.30528E-09	1.8470

West Region

The West Region as defined by the U.S. Census is made up of the Pacific and Mountain Divisions. Data were observed on 41 areas reported in Table 2. Results of the full model indicate that the ACCRA Index and Education Index are significant, but the Area Population and Climate Index are not. The Climate Index had an unexpected positive sign. The best-subsets model results in only the ACCRA Index and Education Index remaining. Thus, the unexpected sign on the Climate Index is not an issue with this model. The best-subsets model has a higher adjusted r^2 than the full model also indicating that the full model should not be used. Neither multicollinearity nor heteroscedasticity is present in either the full model or the best-subsets model.

Table 2: Full Model: West Region Regression				
R Sq = 0.483 Adj. R Sq = 0.426 n = 41				
White's Test 0.054 Breusch-Pagan 0.2263				
Predictor	Coefficients	t Stat	P-value	VIF
Intercept	8174.22	1.080	0.287	NA
ACCRA Index	163.32	1.996	0.054	1.463
Area Population	149.34	0.090	0.929	1.627
Climate Index	6.74	0.188	0.852	1.410
Education Index	95.46	3.365	0.0018	1.615
Best-subsets Model: West Region Regression				
R Sq = 0.482 Adj. R Sq = 0.455 n = 41				
White's Test 0.9370 Breusch-Pagan 0.5861				
Predictor	Coefficients	t Stat	P-value	VIF
Intercept	7815.90	1.081	0.286	NA
ACCRA Index	171.37	2.407	0.021	1.168
Education Index	96.86	4.121	0.0002	1.168

South Region

The South Region as defined by U.S. Census is made up of the West South Central, East South Central and South Atlantic Divisions. Data were observed on 101 areas reported in Table 3. The full model and best-subsets models were identical. Results of the full model indicate that the ACCRA Index, Education Index and Area Population are significant, but the Climate Index is not. Neither multicollinearity nor heteroscedasticity is present. The coefficient of multiple determination is 0.611 for this model. Dropping the Climate Index from the model decreased the adjusted r^2 when compared to the model representing all areas.

Table 3: Full Model: South Region Regression				
R Sq = 0.611 Adj. R Sq = 0.595 n = 101				
White's Test 0.1307 Breusch-Pagan 0.0612				
Predictor	Coefficients	t Stat	P-value	VIF
Intercept	1191.17	0.201	0.841	NA
ACCRA Index	256.71	3.935	0.00016	1.256
Area Population	2619.16	3.152	0.0022	1.824
Climate Index	-25.457	-1.312	0.193	1.196
Education Index	75.195	5.302	7.32779E-07	1.768
Best-subsets Model: South Region Regression				
Same as Full Model				

MIDWEST REGION

The Midwest Region as defined by the U.S. Census is made up of the West North Central and East North Central Divisions. Data were observed on 60 areas reported in Table 4. Results of the full model indicate that the ACCRA Index and Area Population Index are significant, but the Climate Index and Education Index are not. The best-subsets model drops the Education Index, but the Climate Index is still not significant.

The best-subsets model selects the subset model with the highest adjusted r^2 and lowest standard error of the estimate regardless of the significance of the independent variables. Thus, this procedure is concerned primarily with identifying the model that provides the best explanatory power, not with the significance of the independent variables. In this region, the best-subsets model has a higher adjusted r^2 than the full model indicating that the full model should not be used. Neither multicollinearity nor heteroscedasticity is present in either the full model or the best-subsets model.

Table 4: Full Model: Midwest Region Regression				
R Sq = 0.557 Adj. R Sq = 0.5247 n = 60				
White's Test 0.3510 Breusch-Pagan 0.3899				
Predictor	Coefficients	t Stat	P-value	VIF
Intercept	-6555.55	-0.823	0.414	NA
ACCRA Index	376.93	4.608	2.45585E-05	1.367
Area Population	2989.62	3.135	0.003	1.857
Climate Index	-32.66	-1.026	0.310	1.103
Education Index	1.13	0.064	0.949	1.757
Best-subsets Model: Midwest Region Regression				
R Sq = 0.557 Adj. R Sq = 0.533 n = 60				
White's Test 0.5402 Breusch-Pagan 0.2691				
Predictor	Coefficients	t Stat	P-value	VIF
Intercept	-6605.23	-0.841	0.404	NA
ACCRA Index	378.02	4.765	1.37894E-05	1.309
Area Population	3022.00	3.767	0.0004	1.338
Climate Index	-32.48	-1.033	0.306	1.095

NORTHEAST REGION

The Northeast Region as defined by the U.S. Census is made up of the Middle Atlantic and New England Divisions. Data were observed on 17 areas reported in Table 5. Results of the full model indicate that the ACCRA Index and Climate Index are significant, but the Area Population and Education Index are not. The best-subsets model drops the Area Population, and now the ACCRA Index, Climate Index and Education Index are significant. The Climate Index has an unexpected positive sign in the full model and the best-subsets model. This sign may be reflecting the observed higher degree of multicollinearity (note the VIFs) or the effects of unknown correlated independent variables not included in the model. The latter reason is most likely to be the cause since the highest VIFs in the model are less than five. The best-subsets model has a higher adjusted r^2 than the full model indicating that the full model should not be used. Neither multicollinearity nor heteroscedasticity is present in either the full model or the best-subsets model.

Table 5: Full Model: Northeast Region Regression				
R Sq = 0.828 Adj. R Sq = 0.770 n = 17				
White's Test 0.7357 Breusch-Pagan 0.3974				
Predictor	Coefficients	t Stat	P-value	VIF
Intercept	9815.74	2.458	0.030	NA
ACCRA Index	128.48	3.372	0.0055	1.608
Area Population	-1418.96	-0.322	0.753	4.295
Climate Index	130.17	2.216	0.047	1.625
Education Index	98.75	1.465	0.169	4.194
Best-subsets Model: Northeast Region Regression				
R Sq = 0.826 Adj. R Sq = 0.786 n = 17				
White's Test 0.7357 Breusch-Pagan 0.2489				
Predictor	Coefficients	t Stat	P-value	VIF
Intercept	10198.44	2.772	0.016	NA
ACCRA Index	130.56	3.603	0.003	1.562
Climate Index	123.08	2.343	0.036	1.396
Education Index	80.56	2.265	0.041	1.253

CONCLUSIONS

The full model for all areas studied has an r^2 of 0.623 and all independent variables are highly significant. One should not be surprised that the best results are obtained when data from all areas are used in the full model. The all areas data represents the largest sample size from the underlying population compared to any of the regions. This model should most accurately estimate the true unknown population regression. There is no evidence of multicollinearity or heteroscedasticity in this model.

Strengths and weakness begin to appear when the data are disaggregated at the region level. The ACCRA Index is significant in all region models regardless of a full model or best-subsets status. It appears to be the most consistent variable at the regional level in explaining variations in the dependent variable. Our major working hypothesis is that the price level effect captured in the ACCRA Index is the most important independent variable in explaining annual pay regardless of the level of data aggregation. This working hypothesis appears to be upheld. The Area Population dummy variable and the Education Index are significant in two regions out of the four. The Climate Index is significant in only one region and has an unexpected sign in two regions. This index appears to be the least consistent variable at the region level. This result could be expected in light of the procedures used by the *Places Rated Almanac* to generate the index. For example, the *Almanac* ranks Minneapolis and St. Paul in the top 10 when looking at the overall best places to live; however, these cities are near the bottom on the climate index (338 of 354). The low score for climate occurs because of the extreme rather than moderate weather-related conditions in these areas that are then used to compute the index. Regardless of the differences among regions, in every case, the best-subsets models are preferred at the region level based on a higher adjusted r^2 than the full model.

REFERENCES

Breusch, T. and A. Pagan. 1979. A Simple Test for Heteroscedasticity and Random Coefficient Variation, *Econometrica,* 47:1287-1294.

Cherry, Todd L. and Pete T. Tsournos. 2001. Family Ties, Labor Mobility and Interregional Wage Differentials, *The Journal of Regional Analysis and Policy,* 31(1):23-33.

Dumond, Michael J., Berry T. Hirsch and David A. MacPherson. 1999. Wage Differentials Across Labor Markets and Workers: Does Cost of Living Matter?, *Economic Inquiry,* 37(4):577-598.

Eberts, Randall W. and Mark E. Schweitzer. 1994. Regional Wage Convergence and Divergence: Adjusting Wages for Cost-of-Living Differences, *Economic Review (Federal Reserve Bank of Cleveland)*, 30(2):26-37.

Epple, D. and T. Romer. 1991. Mobility and Redistribution, *Journal of Political Economy,* 99:828-858.

Gabriel, Stuart A. and Stuart S. Rosenthal. 1999. Location and the Effect of Demographic Traits on Earning, *Regional Science and Urban Economics,* 29:445-461.

Glaeser, E.L. 1998. Should Transfer Payments be Indexed to Local Price Levels?, *Regional Science and Urban Economics,* 28:1-20.

Montgomery, Edward. 1992. Evidence on Metropolitan Wage Differences Across Industries over Time, *Journal of Urban Economics,* 31:69-83.

Savageau, David with Ralph D'Agostino. 2000. *Places Rated Almanac,* Millennium Edition, IDG Books Worldwide, Inc.

Tiebout, C. 1956. A Pure Theory of Local Public Expenditures, *Journal of Political Economy*, 64:416-424.

White, H. 1980b. A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity, *Econometrica,* 48:817-838.