

ESTIMATING THE COST OF THE CITY WORKER'S FAMILY BUDGET: A MULTIPLE REGRESSION APPROACH

Richard W. Hurd*

School of Labor and Industrial Relations
Michigan State University

The cost of the City Worker's Family Budget (CWFB), computed by the Bureau of Labor Statistics (B. L. S.) for a number of large metropolitan areas, has proved to be a useful measure of intercity differences in the cost of living. This living cost measure has been computed for March 1946, June 1947, October 1949, October 1950, October 1951, Autumn 1959, Autumn 1966, Spring 1967, Spring 1969, and Spring 1970.¹ Unfortunately, CWFB cost figures have been computed for only a limited number of cities. The 1946, 1947, 1949, 1950, and 1951 figures were computed for 34 cities, the 1959 figures for 20 cities, and the 1966, 1967, 1969, and 1970 figures for 39 cities. There are many instances in economic and social research where adjustment of money incomes and money wages to account for intercity variations in the cost of living would be desirable. It would be beneficial if such adjustments were possible for a large number of cities.

The goal of the research reported here is to develop an equation based on the B. L. S. figures which can be used to estimate living costs for other cities in the United States. The detailed data necessary for development of such estimates are available only from the decennial census. Because 1970 Census data were not yet published when this research was undertaken, the most recent year for which the equation would be estimated was 1959. The resulting 1959 CWFB cost estimates should prove especially useful in research employing 1960 Census data. Development of a separate equation for 1970 is planned when 1970 Census data becomes available.

THE 1959 CITY WORKER'S FAMILY BUDGET

The 1959 B. L. S. figures for the CWFB were reported in the Monthly Labor Review in August 1960.² The B. L. S. figures are based on a budget which provides a "modest but adequate" level of living for a family of four: employed husband aged 38, wife not employed outside the house, daughter aged 8, and son aged 13.³ Since the CWFB cost figures estimated below are meant for use as adjusters for money incomes and money wages for purposes of intercity comparisons, the relative costs from city to city are more important than the absolute costs. Thus, the fact that family composition is held constant is more important than the specific composition chosen.

The CWFB is not strictly speaking a price index. Taken into account are variations in consumption patterns as well as variations in prices. "The budget provides a measure of differences in living costs between cities, and not differences in prices only. In addition to differences in price levels, intercity indexes based on the budget reflect climatic or regional differences in the quantity and types of items required to provide the specified standard of living and differences in State and local taxes."⁴ Because of this more flexible approach, the CWFB provides a better measure of intercity variations in the cost of living than would a pure price index.

*This paper was completed while the author was a graduate student at Vanderbilt University. T. Aldrich Finegan and Ben Bolch of Vanderbilt contributed a number of helpful suggestions.

A look at a breakdown of the B. L. S. budget figures into five major categories is instructive (see Table 1). Housing costs, which make up a little less than 20% of the budget, appear to account for much of the intercity variation in living costs. Food costs, making up more than 25% of the budget, vary considerably, though less (both relatively and absolutely) than do housing costs. It is interesting to note that the three southern cities included in the study (Atlanta, Baltimore, and Houston) have the three lowest food costs, resulting in part from different food consumption patterns in the South.⁵ Personal taxes (10% of the budget) also vary markedly, and again are lower in the South -- the three southern cities having three of the four lowest tax levels.⁶ "Other goods and services," though accounting for about 40% of the budget, show less relative variation. "Other costs" (less than 5% of the budget) vary only slightly.

Table 1. Annual Costs of the City Worker's Family Budget -- Autumn 1959

	Total Budget	Housing	Food	Other Goods and Services	Other Costs	Personal Taxes
mean*	6084	1120	1722	2319	262	661
standard deviation*	304	135	116	103	11	70

Source: Helen H. Lamale and Margaret S. Stotz, "The Interim City Worker's Family Budget," Monthly Labor Review, LXXXIII, p. 792.

*Based on figures for 20 cities.

THE BUDGET COST EQUATION

The budget cost equation is an intercity multiple regression equation based on data for the 20 cities covered in the 1959 B. L. S. study. The basic equation is: $C = a + bH + cQ + dP + eS + \epsilon$. As noted above, much of the intercity variation in the cost of the CWFB (C) can be accounted for by variation in housing costs. Thus, an important variable in the budget cost equation is rent (H). Raw rent data, however, reflect not only differences in costs but also differences in housing quality. Thus, a control variable for housing quality (Q) is needed. Another potentially important variable is one accounting for intercity variation in retail prices (P). Also, it was noted that some costs (food and taxes) are lower in the South than elsewhere. Therefore, a regional variable (S) might prove to be significant.⁷ A discussion of the definitions of these variables and the expected sign of the regression coefficient of each follows. The data used in the study are for Standard Metropolitan Statistical Areas (SMSA's), and thus are not restricted by the political boundaries of the cities.

Housing Costs

The B. L. S. shelter budget is based on rent for five room dwellings of specified quality, and includes utility costs.⁸ Available from the 1960 Census of Housing are rent data for each SMSA.⁹ From these data comes the housing cost variable: H is defined as the median gross yearly rent (including utilities) for five room dwellings, based on April 1960 rent levels. A positive relationship between H and C is expected.

Housing Quality

The control chosen for the quality of housing is a housing age variable. Casual observation reveals an obvious relationship between the age of a dwelling and its quality. In general, the newer a building is, the better condition it will be in: though this is not a hard and fast rule, a housing age variable should be a good adjustment for overall housing quality. Define Q as the percent of five room rented dwellings which were more than ten years old in 1960 (as reported in the 1960 Census of Housing¹⁰). A positive relationship with C is expected -- rent held constant, a higher percent of old houses indicates lower overall quality, and thus renters get less for their money.

Retail Prices

Retail prices have an important effect on budget costs. Unfortunately, price data by city are not available. An important determinant of intercity variation in retail prices is undoubtedly intercity variation in the wages paid to retail workers. Since wage data are available, a retail wage variable replaces the proposed retail price variable. The wage variable is intended only as a rough indicator of variations in retail prices, and is used only because no better proxy could be found. Actually, retail wages might prove to be related to C either because they are a primary factor in intercity variation in retail prices (probably the more important reason), or because a high cost of living necessitates higher wage levels in such low paying occupations. Since this research seeks to predict the costs of the CWFB and not to explain them, the direction of causation is not as important as the significance of the relationship.

Data available for each SMSA from the Census of Business¹¹ make calculation of an average retail wage possible. The Census of Business reports the total payroll for retail workers and the total number of retail workers for the workweek ending nearest November 15 during the census year. Total payroll divided by the number of workers gives the average wage of a retail worker for that week. The Census of Business was taken in 1958 and 1963. By computing the average retail wage in each of these years, the 1959 retail wage can be estimated as lying one-fifth of the way between the two computed wage levels. The retail wage variable so computed (W) is expected to have a positive relationship with C .

Region

Food costs and taxes were both lower for southern cities than for other cities in 1959. Thus a variable is needed to express the "southernness" of a SMSA. One possibility is to make S a dummy variable, assigning a value of one to each southern SMSA and a value of zero to all other SMSA's. SMSA's from the following states are designated as southern: Delaware, Maryland, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida, Kentucky, Tennessee, Alabama, Mississippi, Arkansas, Louisiana, Oklahoma, and Texas. The relationship between S and C is expected to be a negative one.

Another variable which captures the "southernness" of a SMSA to some degree is one measuring the racial mix of each SMSA. Define NW as the percent of a SMSA's population which was nonwhite during the 1960 census week. *Ceteris paribus*, one would expect that NW would be negatively related with C because in a SMSA with a large nonwhite population there would be available a large supply of low wage service workers (domestic servants, yard workers, laundry workers, dish washers, cooks, hospital workers, etc.) providing a downward force on living costs.¹² Additionally, the readily observable relationship between the racial mix of a SMSA and its geographical

location -- most SMSA's with a high percent of nonwhites are located in the South¹³ -- implies a negative relationship between NW and C.

Because of the high correlation between S and NW, they cannot be used together in the budget cost equation. NW seems superior because it serves as a rough measure of "southernness" and at the same time reflects market forces which affect living costs. Dummy variables such as S are not totally satisfactory because of their all-or-nothing character; there are clearly degrees of "southernness," and these degrees cannot be captured by the regional dummy. For these reasons, NW appears in the budget cost equation used below for the cost estimates of the CWFB. The estimated equation using S in place of NW is also reported.

RESULTS

Based on the testing, two versions of the budget equation seem to be successful estimators of the cost of the CWFB.

$$I. C = a_1 + b_1H + c_1Q + d_1W + e_1NW + \epsilon_1$$

$$II. C = a_2 + b_2H + c_2Q + d_2W + e_2S + \epsilon_2$$

C = cost of the CWFB, 1959 (based on November 1959 prices)

H = median gross yearly rent for a five room dwelling (based on April 1960 rent levels)

Q = the percent of five room rented dwellings occupied in April 1960 which were built before 1950

W = mean yearly retail wage (based on estimated November 1959 wages)

NW = the percent of population which was nonwhite in April 1960

S = 1 for all southern cities, 0 for all other cities.

The estimate for equation I is:

$$I. C = 2086 + 1.73 H + 15.1 Q + 0.37 W - 14.7 NW$$

(2.56)	(5.91)	(2.68)	(2.29)	(-2.46)
**	***	**	**	**

Given in parentheses beneath each regression coefficient is its t-value; the significance of each t-value is indicated by asterisks; *** = 99%, ** = 95%, * = 90%. The multiple R^2 for equation I, adjusted for degrees of freedom, is .81.

The estimate for equation II is:

$$II. C = 2571 + 1.36 H + 13.8 Q + 0.33 W - 312 S$$

(2.90)	(4.96)	(2.47)	(2.03)	(-2.73)
**	***	**	*	**

The multiple R^2 for equation II, adjusted for degrees of freedom, is .82.¹⁴

Both equations seem to be successful estimators. -- all regression coefficients have the expected sign and each is significant at the 90% level or higher. Equation I is chosen as the estimating equation for the reasons pre-

viously mentioned. With a multiple coefficient of determination of .81 and with four significant variables, use of equation I as a budget cost estimator seems to be justified.

Cost estimates based on equation I may be obtained from the author. Estimates are available for the 20 SMSA's included in the 1959 B.L.S. study and for the 81 additional SMSA's with a 1960 population of 250,000 or greater. The equation could be used for estimating 1959 CWFB costs for smaller SMSA's (data are available for all SMSA's with a minimum population of 100,000), but it should be pointed out that the smallest SMSA in the sample on which the equation is based in Scranton, Pa., with a 1960 population of 234,000.

A comparison of my 1959 estimates with the 1966 B.L.S. figures is interesting. Of the 20 cities included in the 1959 B.L.S. study, 18 were also included in the 1966 B.L.S. study. My 1959 estimates include these 18 cities plus an additional 13 of the cities included in the 1966 B.L.S. study. Although the CWFB cost levels for these cities may not prove to be highly correlated between 1959 and 1966, one would expect that the ranking of the cities according to budget costs would not change substantially over the seven year period. To get some indication of the validity of my 1959 estimates, the following test was designed.

First, employing the Spearman rank correlation coefficient, the correlation between the 1959 B.L.S. ranking and the 1966 B.L.S. ranking for the 18 cities included in both studies is calculated (r_{s1}). Then, again employing the Spearman coefficient, the correlation between the ranking of my 1959 estimates and the ranking of the 1966 B.L.S. figures for the same 18 cities is calculated (r_{s2}). Finally, the correlation between the ranking of my 1959 estimates and the ranking of the 1966 B.L.S. figures for all of the 31 cities included in both studies is calculated (r_{s3}). The coefficients r_{s1} and r_{s2} should be approximately the same. If they are, and if r_{s3} is not significantly smaller than r_{s1} , the conclusion follows that my 1959 estimates are related to each other in the approximate rank order expected from the 1966 B.L.S. study; use of equation I to make budget cost estimates for cities not included in the 1959 B.L.S. study would gain a degree of validation.

The results of this test are encouraging. First, r_{s1} was calculated to be .64, significant at the 99% level. Next, r_{s2} was also calculated to be .64. Finally, r_{s3} was calculated to be .71, also significant at the 99% level. Thus, r_{s3} actually turned out to be higher than expected, an indication that for those cities where comparison is possible, my 1959 estimates are reasonably accurate.¹⁵

The CWFB cost estimates based on equation I can be used as a guide to approximate living costs in 1959, but are certainly not as accurate as the computed B.L.S. figures. They should prove useful, however, since living costs estimates for cities not included in the 1959 B.L.S. sample have heretofore been unavailable. The equation and the resulting CWFB cost estimates are for 1959 only. A separate equation would be needed to make estimates for other years.

FOOTNOTES

¹1946 and 1947: Monthly Labor Review, February 1948, pp. 133-170; 1949 and 1950: Monthly Labor Review, February 1951, pp. 152-155; 1951: Monthly Labor Review, May 1952, pp. 520-522; 1959: Monthly Labor Review, August 1960, pp. 785-808; 1966: Monthly Labor Review, November 1967, pp. 1-8; 1967: Monthly Labor Review, April 1969, pp. 3-16; 1969 and 1970: available from the Bureau of Labor Statistics.

²Helen H. Lamale and Margaret S. Stotz, "The Interim City Worker's Family Budget," Monthly Labor Review, LXXXIII, pp. 785-808.

³ibid., pp. 785-786.

⁴ibid., p. 808.

⁵ibid., p. 787

⁶ibid., p. 792

⁷A variable for city size was also included in earlier testing in an attempt to detect a relationship between living costs and population. However, no significant relationship was revealed for 1959. It is possible that the insignificance of the population variable resulted from the biased nature of the 20 city sample -- only one of the cities (Scranton, Pa.) had a 1960 population of less than 800,000. In a trail equation run with 1966 B.L.S. budget cost figures (available for 39 cities including several smaller ones), population was positively related with the cost of living, and this relationship was significant at the 95% level.

⁸Lamale and Stotz, p. 792.

⁹U.S. Department of Commerce, Bureau of the Census, 1960 Census of Housing, Volume II, Metropolitan Housing.

¹⁰ibid.

¹¹U.S. Department of Commerce, Bureau of the Census, United States Census of Business: 1958, Volume II, Retail Trade Area Statistics, and 1963 Census of Business, Volume II, Retail Trade Area Statistics.

¹²This force was not accounted for in W above, since racial discrimination has for the most part excluded nonwhites from the labor market for retail clerks.

¹³The demographic trend of a northward migration of blacks together with the movement of whites to the suburbs has markedly increased the non-white percent of population within city limits in many northern cities, but the percentage effect on SMSA's as a whole has been much smaller.

¹⁴With both NW and S included the following equation results:

$$C = 3037 + 1.55 H + 10.3 Q + 0.24 W - 239 S - 10.4 NW$$

(3.50)	(5.60)	(1.85)	(1.51)	(-2.09)	(-1.79)
***	***	*		*	*

The decreased significance of all variables except H probably results from the presence of multicollinearity: $r_{S,NW} = .58$, $r_{S,W} = -.46$, $r_{Q,NW} = -.55$.

¹⁵As it turned out, the simple correlation coefficients using the traditional least squares method were quite similar: r_1 (corresponding to r_{s_1}) was .65, r_2 (corresponding to r_{s_2}) was .67, and r_3 (corresponding to r_{s_3}) was .74.