Title

Cost of Living and Income by Urban Size

Authors

Alonsi, William

Fajans, Michael

Publication Date

1970-06-01

Working Paper 128

Cost of Living and Income by Urban Size

William Alonso and Michael Fajans

June 1970

University of California at Berkeley
$9.00

Institute of Urban & Regional Development
University of California, Berkeley

COST OF LIVING AND INCOME BY URBAN SIZE

William Alonso

and

Michael Fajans

Department of City and Regional Planning
University of California, Berkeley
July 1970
Working Paper No. 128

This paper was supported by a grant from the Economic Developement Administration.

It is generally believed that it is more expensive to live in large cities than in small ones. The matter is of obvious interest to most people, partly because such knowledge may help them if they are thinking of moving and partly because people enjoy comparing their own circumstances with those of others. It is also of interest to large corporations whose frequent practice is to make cost of living adjustments according to local costs, and to industry in general in their location decisions as background information for wages and other costs. And it is of interest to national policy, which in the United States increasingly follows the example of other countries in its concern for the productivity and levels of welfare of cities of different sizes. Finally, it should be of considerable interest to economists, for variations in prices among cities are central to an understanding both of the externalities of urban agglomeration and of the degree of factor integration, and thus development, of the national economy.

Our central concern is with national policy. In brief, if the urban areas of a country are each viewed as an aggregate productive unit, and if the relation of inputs to outputs vary systematically with urban size, then national policy should try to bring about the most efficient sizes in the context of objectives of national growth. At present our knowledge as to these relations may be summarized as follows: (1) Measures of output would ideally consist of Net Regional Income per capita for urban areas. This information is not directly available for the United States, but data for West Germany [3] and for Japan [7] show a strong rise in per capita Gross Regional Product with urban size. In almost every country there is a strong rise in per capita income with urban size, and some of the findings on this will be presented below. (2) Measures of inputs are less developed, but three principal types may be identified: (a) Municipal and public infrastructure costs have been frequently studied (see [6] for a recent overview), and in general find a U-shaped curve with a bottom in the 50,000 to 250,000 population range. However, the association is relatively weak, and population tends to lose significance under multivariate approaches. Moreover, the range of these costs is considerably smaller than the range in increases in product with urban size. (b) There have been no studies of the variations in producers' costs with urban size in this country. The most detailed study has been for a limited range of urban sizes in India [8], and this found a small decline of such costs with increasing urban size. (c) Variations in consumer costs have not been studied, and this report seems to be the first systematic analysis. They would be of considerable importance to determine whether the rise in per capita income with urban size is maintained after the figures are adjusted for differences in cost of living. With some qualifications, we find that they are.

Summary of Findings

The cost of living, as measured by the Bureau of Labor Statistics [4] is only weakly associated with urban size, especially for lower incomes. Other factors are more important, among them local climate and percent non-white. But the strongest association is with local income. Since the construction of the B.L.S. figures appears to have made considerable allowance for local consumption patterns, the higher income in larger urban areas and the greater income elasticity of higher forms of consumption would lead to an apparent rise in the cost of living in areas of higher income; but this rise would be one in the standard rather than the cost of living, and would reflect demand rather than supply conditions. In view of this, since the statistical association of the B.L.S. Budgets to income is far stronger than that to urban size, and since under step-wise multiple regression local income dominates and urban size fails to enter significantly, we conclude that it is not more expensive to live in larger urban areas. Deflation of local income by the B.L.S. Budget figures did not reverse the positive association of income and urban size. However, a recent opinion poll asking people what a modest but sufficient income would be in that community shows a far sharper rise than do the B.L.S. figures, showing that expectations rise faster than income with urban size. Although the findings are statistically extremely weak, and theoretically doubtful, it is possible that the relation of size to local income becomes negative for extremely large urban sizes.

Factor analyses of the data, on the whole, emphasize regional factors (principally the distinction of the South from the rest of the country, and secondarily a Western factor) rather than the factor of population magnitude. Regional factors correlate, on the whole, most strongly with both B.L.S. Budget and income indicating, within the fragility of the statistical techniques, that the American economy is imperfectly integrated.

In short, we find that consumer costs rise slightly with urban size, but not enough to offset the rise in income with urban size. The common belief that bigger places are more costly appears the result of higher expectations rather than higher prices.

The Data and the Variables

The principal dependent variable, the 1967 B.L.S. City Worker's Family Budget applies to a family of four, consisting of an employed father 38 years old, his wife, a boy of thirteen, and a girl of eight. Three levels of consumption are considered: a low one averaging $6,000, a moderate one averaging $9,000, and a higher one averaging $13,000. These Budgets are not simple cost of living figures, where a constant market basket is kept constant while prices are allowed to vary; rather, with the exception of medical costs, the elements in the Budget vary by type, quantity, and price from city to city, taking into account such factors as local variations in climate and local life-styles. This data is available for 38 metropolitan areas ranging in population from slightly over 100,000 to over 10,000,000. Data from a few other sources, less comprehensive or less reliable, were also considered in our work.

Two other variables deserve mention because they may be less familiar to readers. Population potential is a variable constructed by the formula $V_i = \sum_{j=1}^{n} P_j/d_{ij}$, where d_{ij} is the distance between i and j. This variable may be interpreted as a measure of the accessibility to the rest of the national population available to a resident of city _i_. It is a variable commonly found useful and statistically significant in the analysis of territorial data. It was calculated for the central county of each metropolitan area on the basis of the 1960 population of all of the U.S. counties. Because of the difficulty of defining satisfactorily d_{ii}, we used Warntz's suggested adjustment, which assumes the city to be a circle with a homogeneous population density. However, we found that this approach (which is an approximation of the integral over area of density divided by distance) was highly sensitive to the relative areal and population size of the central metropolitan county, and that this introduced a large relative measurement error. It is possible that this imprecision in measurement has undermined the significance of this variable.

The other variable that requires explaining is Degree Days. This is a measure of the coldness of a place, operationally defined as the number of degrees below 65°F times the number of days when such a temperature occurred.

The notation for the principle variables in our analysis is as follows:

$B_{l,m,h}$: City Worker's Family Budget, 1967; the subscript denotes low, medium, and high budget;

the components of the budget are:

F: food;

H: housing;

T: transportation;

C: clothing;

M: medical;

X: personal tax.

Y: mean per capita income, Standard Metropolitan Statistical Area, 1960.

P: population of Standard Metropolitan Statistical Area, according to the 1960 Census or the 1966 Census estimates, as appropriate. Population is in thousands.

lnP: natural logarithm of P.

V: population potential, as defined above.

D: degree days, as defined above.

N: percent of the population which in non-white.

G: percent population growth, 1960-66.

s: level of significance of the t-value of the coefficient.

R^2: coefficient of multiple determination adjusted for degrees of freedom.

Budget

A slight positive association emerges if population or its logarithm is regressed simply against Budget. The best equations of this set are:

$B_l = 5340 + 83.6 \ln P$ $\quad R^2 = .08 \; s > .05$

$B_m = 8870 + .0922 P$ $\quad R^2 = .15 \; s > .01$

$B_h = 12600 + .177 P$ $\quad R^2 = .27 \; s > .001$

The association is greater for higher incomes. For a population range of 1,000,000 to 10,000,000, the difference would amount to \$192 (or 3% or the central figure) for the low budget, and to $1,590 (or 12% of the high). The association is slight, of course, amounting to less than 28% of the variance of the Budget in the best case, but it is reinforced by other evidence. An earlier B.L.S. [11] set of budgets has a Spearman rank correlation coefficient of .45 with the metropolitan populations which is not quite significant at the .01 level. Budgets from a private study [9] for families at various income levels exhibit Spearman coefficients of .52 for families at the $6,000 level and progressively lower coefficients at higher incomes. At the $24,000 level, the coefficient is .42, which is not significant at the .01 level. The range from least to most expensive was 20% of the average for the lowest income, and 18% for the highest. It should be noted that this data reverses the pattern of increasing significance of urban size for higher incomes which appears in our principal B.L.S. data. A further study [1] shows a very strong Spearman coefficient of .98, but the range is only 11%.

The coefficients of population or its log remain quite stable, rising slightly, in a step-wise multiple regression, and their level of significance rises considerably. The variables offered were P, lnP, V, D, G, and N. Cutting off at the .1 level of significance and omitting the non-significant variables, the following equations result:

$$B_l = 5440 - 20.4\ N + 104\ \ln P \qquad R^2 = .35 \qquad s_N > .001,\ s_{\ln P} > .01$$

$$B_m = 8640 + .0945\ D + .0993\ P - 19.7\ N \qquad R^2 = .50 \qquad s_D > .01,\ s_P > .001,\ s_N > .05$$

$$B_h = 13040 + .188\ P - 37.5\ N \qquad R^2 = .43 \qquad s_P > .0001,\ s_n > .01$$

The appearance of degree days (D) in the medium budget is not surprising, of course, since it may be expected to relate both to clothing costs and to housing costs through heating (see Appendix). The appearance of percent of population which is non-white (N) is harder to interpret in itself, but we shall see below that its significance is probably regional, associated with lower costs in the South in general.

But by far the variable most strongly associated with budget is income. Regressing budgets against all variables, with and without income, and against income alone, one obtains:

	R^2 with all variables	R^2 with all but income	R^2 with income alone
B_1	.64	.35	.56
B_m	.72	.50	.52
B_h	.65	.43	.46

This raises an important question as to what degree these Budgets represent costs of living and to what degree they represent levels of living. The descriptive material in the B.L.S. publication make it clear that there is allowed considerable inter-city variation by type and quantity of consumption, as well as price. It suggests that the implied adjustments in the construction of the local market basket, based on observed behavior and judgment, may in fact be a reflection of income and of income-elasticities for superior goods rather than of local prices and conditions. At any rate, when income is allowed in the step-wise regressions, neither population nor its log appear at significant levels.

A recent Gallup Poll [5] sheds light on a related aspect. It asked "What is the smallest amount of money a family of four (husband, wife, and two children) needs each week to get along in this community?" This is equivalent in intent to the B.L.S. low budget, and a comparison of the

Gallup Poll responses with the results of the regression of the B.L.S. data with lnP (converted to weekly figures) yields:

Community Size	Gallup Poll (1970)	B.L.S. Low Budget (1967)
1,000,000 or over	149	115-120
500,000 to 999,999	148	113
50,000 to 499,999	104	110
2,500 to 49,999	101	107
under 2,500	101	-

If we interpret the Gallup figures as subjective estimates, based on expectations as much as on cost differences, and the B.L.S. figures as relatively more objective measures of cost and need differences, we have a plausible explanation of the widespread belief that big cities are more expensive. It is not so much that things cost more there as that people expect more; having the best, or even the good, where the range of options is very large, is far more expensive. In brief, one can probably live for about the same in big cities as in small, but one is less content to do so. The same point is made by an experienced location consultant in discussing the movement of a firm to New York. He predicted higher wage costs on the basis of the "expenditure standard of a community", and explained "While the cost of living, measured by the market-basket yardstick, is but little higher here. . . this ignores subtle factors known to all who have lived in New York. Its tempo of expenditure is vastly higher. There is an immense choice in how you spend here, and unceasing incitement to spend. It takes more character to be thrifty. In smaller places there is a saner regard for the potential worth of a dollar. Pleasures and luxuries are simpler, less expensive, probably less frequent. In New York the wage earner demands more money principally because the family spends more." [13, p. 69].

Income

Since family budgets are somewhat associated with population size, and strongly associated with income, it is useful to look at the relation of income to urban size. For the same 38 metropoles for which we have the B.L.S. Budgets, the relation is:

$$Y = 4110 + 321 \ln P \qquad R^2 = .34 \qquad s > .0001$$

Although the regression accounts for only one third of the variance in metropolitan incomes, this is considerable since we omit the many other variables such as local resources and socio-economic circumstances that are at play. At any rate, the R^2 is greater than that of the comparable regression on budget, and the coefficient of lnP is nearly four times as great, in spite of a 12.8% increase in the cost of living from 1960 (which is the date of the income figures) to 1967. Adjusting the income equation by the increase in the Consumer Price Index, we see clearly that income rises faster with population than costs as measured in the family budgets:

Population	Calculated Low Budget	Calculated Income (adjusted)	Ratio Y/B_1
100,000	5730	6300	1.10
1,000,000	5920	7130	1.20
10,000,000	6110	7960	1.30

Further confirmation is obtained by the positive coefficients in a regression of lnP upon a rough index of level of living, constructed by dividing budget into income:

$$Y/B_1 = .795 + .0395 \ln P \qquad R^2 = .37 \qquad s > .0001$$

$$Y/B_m = .535 + .0238 \ln P \qquad R^2 = .30 \qquad s > .001$$

Because of the possibility that the relation of income to urban size may not be monotonic, income and income deflated by budget were run against population and its logarithm, resulting in the following equations:

$$Y = 3740 + 386 \ln P - .0410 P \qquad R^2 = .33 \qquad s_{\ln P} > .01$$

$$s_P = .52$$

$$Y/B_1 = 1722 + .0521 \ln P - .803(10^{-5}) P \qquad R^2 = .37 \qquad s_{\ln P} > .001$$

$$s_P = .28$$

$$Y/B_m = .442 + .0400 \ln P - .102(10^{-4})\, P \qquad R^2 = .36 \qquad s_{\ln P} > .001$$
$$s_P > .1$$

The coefficients of multiple determination are hardly changed by adding a second transformation of population, but all three versions do show a downturn, at a population of 9,400,000 for income, and at populations of 6,400,000 and 3,900,000 for income deflated by the low and moderate budgets respectively. Such a finding could be of great potential significance, since it argues some sort of maximum size beyond which diminishing returns set in. However, this result is extremely weak for several reasons: (a) The apparent turning point is extremely unreliable; using the standard errors of the coefficients (see [2] for the technique), in the case of income the downturn at 9,400,000 has a standard error of 13,000,000; in the case of income deflated by the low budget, the point of inflexion at 6,400,000 has a standard error of 6,100,000. (b) It appears possible that the areal definition of the largest metropolitan areas excludes a number of high income exurbs, so that the actual mean income of these metropoles is understated in the Census, and the apparent downturn is not a real one. (c) A lower per capita income in the largest metropoles may be consistent with overall efficiency for the system of cities since these cities perform systemic functions, as ports of entry for unacculturated, low-income migrants, and as seed beds for innovative infant industries which, being in the process of rapid change, have lower capital investment per worker and therefore lower marginal output and wages per worker (see [3], [10], [12]). As migrants and industries mature, they move to secondary centers where they contribute to higher incomes.

Factor Analysis

The multiple regression analysis suffered from intercorrelation among the variables, which resulted in considerable uncertainty in the coefficients. Examination of some of these correlations is enlightening if not surprising. For instance, percentage non-white was negatively associated with income (especially central city income); degree days are negatively associated with population growth in 1960-66 (cities in warm climates are growing faster); degree days are positively associated with income, in a puritanical fashion. But the web of interconnections is such that factor analyzing the data seemed a useful way of sorting things out. As will be seen below, the principal outcome of the exercise was the uncovering of strong regional effects (South and West), stronger perhaps than population effects.

In the first instance, the following variables were subjected to a Varimax rotation: Y, P, lnP, V, G, D, and N. Two factors emerged which accounted for 73% of the commonality. The stronger factor, which we call Magnitude, has high correlations with P, lnP, and V, and a moderate one with Y. The second proves to be a regional factor, with high positive correlation with D, and high negative correlations with N and G. With the exception of Los Angeles, the ten cities with lowest scores in this factor were in the South, and therefore we have labelled it Non-South. We then regressed these factors on the various budgets and their components. Table 1 shows their correlation with these two variables. The most striking finding is the preponderance of the Non-South factor. Magnitude enters first and significantly only in the High Budget and in Housing for the High Budget.

TABLE 1

Correlation Low, Medium, and High Budgets, and their Components
with the factors Magnitude and Non-South

Dependent Variable	Low Budget	Moderate Budget	High Budget
Budget	R^2 = .41 NS, M	R^2 = .61 NS, M	R^2 = .55 M, NS
Food	R^2 = .50 NS, M	R^2 = .44 NS, M	R^2 = .50 NS, M
Housing	R^2 = ml NS*, M*	R^2 = .48 NS, M	R^2 = .42 M, NS
Transportation	R^2 = ml (-)M*, NS*	R^2 = ml (-)M*, NS*	R^2 = ml M*, (-)NS*
Clothing	R^2 = .55 NS, M	R^2 = .52 NS, M	R^2 = .30 NS, M
Medical	R^2 = ml M*, (-)NS*	R^2 = ml M*, NS*	R^2 = ml M*, (-)NS*
Personal Tax	R^2 = .56 NS, M	R^2 = .50 NS, M	R^2 = .38 NS, M

* = not significane at .01 level

(-) = negative coefficient

Factors entered in order of importance

The second analysis reserved income as the independent variable and factor analyzed P, lnP, G, V, D, N, and, for each budget level, B, F, H, T, C, M, and X. For low budget, four factors emerged, accounting for 82% of the commonality: (1) a South factor, similar to the Non-South above but with the signs reversed; (2) a Magnitude factor; (3) a Housing-Budget factor, strongly correlated with these variables, and slightly less with transportation and personal taxes; and (4) a West factor, correlated to good weather (-D) and high medical costs, with which the six Western metropoles correlated much more than the others. The factors were similar for the moderate and high budgets, except that Magnitude and Housing-Budget were compressed into a single second factor, so that their factors were Non-South, Magnitude and West. The regressions of these factors on income are shown in Table 2, but the results are inconclusive, except for the importance of the South (or Non-South) variable. The variable Magnitude, although showing a positive sign, barely fails in each case to qualify at the .01 level of significance. Except for the equivocal positive relation to Housing-Budget for Low Budget, the South (or Non-South) variable dominates. The other regional variable, West, barely misses our criterion of significance for moderate and high budget.

TABLE 2

Correlation of Income with the Factors of Magnitude (and Housing-Budget for Low Budget), South (or Non-South) and West

Dependent Variable	Low Budget	Moderate Budget	High Budget
Income	R^2=.56 H-B,(-)S,M*,W*	R^2=.46 NS, W*, M*	R^2=.42 NS, M*, W*

* = not significant at .01 level

(-) = Negative coefficient

Factors entered in order of importance

These results are somewhat at variance with those of the regression analysis, and to much existing literature which shows strong correlation of urban size and income. The factor analysis findings seem to say that the most important factor is whether or not a metropolis is in the South, and that scale and Westerness, while also important, take a back seat. The data and the techniques are sufficiently fragile that no single finding must be interpreted as establishing a fact. Our interpretation is that, as shown by our regression and by other works, income is to some extent a function of urban size, but that the lack of integration of the American economy is reflected in the strong regional flavor of the regressions of income upon the factors.

APPENDIX

Principal findings for income and the components of budget when regressed upon all other variables

Income

For the 38 metropoles included in the B.L.S. study, the step-wise multiple regression found income a function of lnP, and of D at very high levels of significance, with V appearing with a negative coefficient with a significance level of .08 with $R^2 = .54$. In other experiments this last variable, which was expected to be positive since it measures access to the national population, also scored negatively. This is counter to theory and much associated empirical evidence, and may be the result of the intercorrelation of variables. When income per capita was regressed logarithmically in a pseudo Cobb-Douglas function on Y and V for 211 metropolitan areas, the result was $Y = 8.52P^{.0831}V^{.0856}$, with $R^2 = .19$, but high significance for the coefficients. Parallel research by the senior author and O. Fisch has established $Y = 5.01\ P^{.0661}V^{.0866}$, with a much higher $R^2 - .51$. The principal differences in the latter study (aside from the population being measured in full rather than in thousands) are that the potential is income potential $(\Sigma_j P_j Y_j / d_{ij})$ measured over the metropolitan areas rather than all counties, and that self-potential is excluded. The similarity of the exponents is striking, and the improvement of the R^2 may be the result of the elimination of noise in the V variable through a better definition.

Budget Components

Food. As expected, food is non-elastic with respect to income, constituting 28% of the low budget but only 20% of the high. Y is a strong predictor for high and low budgets, but not very strong for medium. For B_1 regression yields, in order of appearance, association with G (negative), P, D, N (negative), lnP, V, for an R^2 = .52, meaning that food is cheapest in slow-growing, large, cold, white cities which are in high potential areas. However, regression on Y, N (negative) and V yields an R^2 of .60. The strong relation with income raises again the question of the meaning of the B.L.S. figures. From the point of view of theory, the significance and positive coefficients of P and V accord with rent and central place theory, but their effects are small.

Housing. The association with population is weak, in contrast with what might be expected from rent theory. The association is slightly higher for higher budgets. It is most strongly associated with income (R^2 = .31, .42, .35 for low, moderate and high budgets). Aside from income, the principal variables with which it is associated are lnP, D, and V in the case of low budget. For medium budget, the strongest associations are with D, and P. V accounts for a high percent of the variance in the high budget, which makes theoretical sense on the basis of the suburban character of high income housing and the greater sensitivity of suburban land to agricultural land values as marginal rent.

Transportation. Conventional wisdom is certain that transportation costs rise with urban size and its consequent congestion. But there is a surprise here in the B.L.S. data. For the low budget, the regression equation is $T_1 = 310 - .011\ P + 20.2\ \ln P$, $R^2 = .39$. This function rises at first, but then turns down. The significance of the coefficients is extremely high, and the function rises to a population of 1,840,000 and declines thereafter. The turndown for the moderate budget occurs at 940,000. For the high budget the function increases monotonically (both coefficients are positive), but the significance of the coefficients and the R^2 are nil.

Clothing. Quite sensibly, clothing is correlated with degree days (with correlations of .27, .29, and .23 for the three levels). When other variables are entered, the principal (although not strong) effects are positive with income, negative with non-white and, curiously, negative with population growth. This last may be related with the more rapid growth of cities in warm areas.

Medical. All three budget levels varied consistently: positively with income, negatively with degree days (curiously, cold cities are cheaper), and positively with population.

Personal taxes. These relate primarily with degree days and with income, this last being the stronger variable for high budget. While snow removal and the like make it sensible that colder cities should have higher costs and therefore higher taxes, the strong association with income suggests an income-elasticity effect in the demand for public services rather than a cost effect.

William Alonso and Michael Fajans, COST OF LIVING AND INCOME BY URBAN SIZE, Working Paper No. 128, Institute of Urban and Regional Development, University of California, Berkeley.

FOOTNOTES AND BIBLIOGRAPHY

1. A.I.R.I.N.C., "Index of Cost of Commodities and Services," unpublished, Cambridge, 1966.

2. Alonso, William, "Predicting Best with Imperfect Data," Journal of the American Institute of Planners, July 1968.

3. Alonso, William, "Urban and Regional Imbalances in Economic Development," Economic Development and Cultural Change, October 1968.

4. Bureau of Labor Statistics, U.S. Department of Labor, "Three Standards of Living for an Urban Family of Four Persons," 1969.

5. Gallup, G., "Typical U.S. Family of Four Needs $120 per Week to Make Ends Meet," American Institute of Public Opinion, Princeton, N.Y., 1970.

6. Hirsch, W., "The Supply of Urban Public Services" in Issues in Urban Economics (H. Perloff and L. Wingo, eds.), Johns Hopkins Press, Baltimore, 1968.

7. Mera, K., "On the Concentration of Urbanization and Economic Efficiency," International Bank for Reconstruction and Development, Economics Department Working Paper 74, 1970.

8. Morse, Richard, et. al., "Costs of Urban Infrastructure for Industry as Related to City Size in Developing Countries: India Case Study," Stanford Research Institute, Menlo Park, California, 1968.

9. National Industrial Conference Board, Inc., "Comparative Living Costs in 20 Cities," Technical Paper No. 14, New York, 1964.

10. Pred, A. R., The Spatial Dynamics of U.S. Urban-Industrial Growth, 1800-1914, M.I.T. Press, Cambridge, 1966.

11. Statistical Abstract of the United States, 1965; Table 499.

12. Vernon, R., Metropolis 1985, Doubleday Anchor, New York, 1963.

13. Yaseen, L. C., Plant Location, American Research Council, New York, 1960.