John D. Kasarda

A. THE IMPACT OF SUBURBAN POPULATION GROWTH ON CENTRAL CITY SERVICE FUNCTIONS

Since World War II, dynamic yet disproportionate growth has been occurring in our metropolitan areas. Between 1950 and 1970, the standard metropolitan statistical areas (SMSAs) accounted for approximately 80% of the national increase in population. Almost all the metropolitan growth, however, has occurred in the suburban rings surrounding the central cities, while most central cities have grown very little and many have experienced a decline in population. As a result, for SMSAs as a whole, more people are now residing in the suburban rings than in the central cities.

Concurrent with this population redistribution has been a significant alteration in the social morphology of our urban areas. The compact urban community of nineteenth-century America has been replaced by the diffuse metropolitan area as entire communities have become territorially specialized and dependent on one another. This new entity consists of a large central city nucleus and a plethora of politically autonomous but functionally dependent suburban populations serviced by the central city. As a specialized service center, the city's facilities are utilized intensely by a large part of the suburban population for employment, shopping, recreation, professional services, and other needs.

This study examines the central city as a service center for the suburban population. More specifically, it addresses two issues: (1) does the size of the suburban population have a significant impact on service functions performed in the central city, and (2) what effect, if any, have recent increases in the suburban population had on public services provided by central city governments? Primary attention is given to the latter issue and its implications for central city planning and metropolitan reorganization.

Source: John D. Kasarda, "The Impact of Suburban Population Growth on Central City Service Functions," *American Journal of Sociology*, 1972, *77*, 1111-1124.

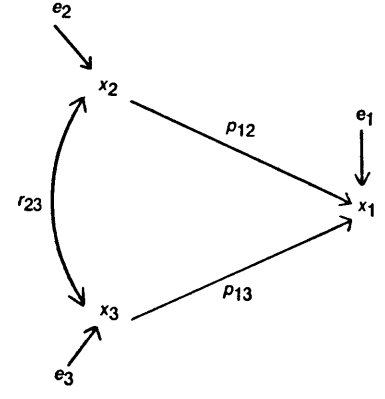
I wish to thank Amos Hawley and H. M. Blalock, Jr. for their useful comments on an earlier draft of this manuscript. Any shortcomings, of course, are my responsibility.

DATA AND METHOD

The metropolitan communities to be examined are all SMSAs as of 1950 ($N = 168$). Data representing four broad, but distinct, categories of service functions performed in central cities were obtained from the 1948, 1958,and 1967 Censuses of Business and the 1950, 1960, and 1968-69 Compendia of City Government Finances (U.S. Bureau of the Census). The categories are: (1) retail trade, (2) wholesale trade, (3) business and repair services, and (4) public services provided by central city governments. Sales and receipts for central city retail trade, wholesale trade, and business and repair services are used as indicators of the magnitude of these three categories of service functions, whereas annual operating expenditures for noneducational services provided by central city governments are used as indicators of the magnitude of the public service functions. For cross-sectional analysis, the sales, receipts, and operating expenditures are expressed in terms of "amount per central city resident"; for longitudinal analysis, these data are converted to constant dollars and expressed in terms of first differences.

Demographic data were obtained from the 1950, 1960, and 1970 censuses of population. The suburban population is defined as all population residing outside the central city but within the metropolitan area for those SMSAs that contain a single central city, and as all population residing

FIGURE 1



Key: X_1, central city service variable; X_2, central city population size; X_3, suburban population size; p_{12} and p_{13}, path coefficients; r_{23}, zero-order correlation coefficient; e_1, e_2, and e_3, error terms.

outside the largest central city but within the metropolitan area of SMSAs that contain more than one central city.[1]

The primary method used is path analysis, which provides an algorithm for decomposing the total correlation between the dependent and independent variables so that the direct effects of each independent variable on the dependent variables may be ascertained. For example, if central city service functions vary with both the size of the central city population and the size of the suburban population, then a path model mapping the relationship appears as shown in Figure 1.

In a multivariate path model such as this, the zero-order correlation between an independent variable and the dependent variable is the sum of its direct effect via the path from that independent variable to the dependent variable and its *indirect effect*[2] through its correlation with the other independent variable (Land 1969, pp. 12-15). In terms of the coefficients,

$$r_{12} = p_{12} + r_{23}\,p_{13},$$
$$r_{13} = p_{13} + r_{23}\,p_{12}.$$

The path coefficients (p_{12} and p_{13}) measure the variance in X_1 for which each independent variable (X_2 or X_3) is *directly* responsible, while the other independent variable is held constant.[3] Therefore, by computing path coefficients, we will be able to determine the direct effect of variation (or change) in both the central city and suburban populations on each central city service function as well as the direct effects of other independent variables.

TABLE 1

Path Coefficients between per Capita Measures of Central City Service Functions and Population Size (Log) of Central Cities and Suburban Areas, 1960 and 1970

	Central City Service Functions			
Population Size	*Retail Trade (1958)*	*Wholesale Trade (1958)*	*Business and Repair Services (1958)*	*Public Services (1960)*
1960:				
Central city	—.76*	.33*	.11	.01
Suburban	.83*	.27*	.54*	.39*
Population Size	*Retail Trade (1967)*	*Wholesale Trade (1967)*	*Business and Repair Services (1967)*	*Public Services (1969)*
1970:				
Central city	—.83*	.16	.04	.06
Suburban area	.61*	.36*	.53*	.41*

* Significant at .001.

1. Since much of the detailed data analyzed in this study were available for only the largest central city of those SMSAs that contain two or more central cities, the largest city was considered the nucleus, and the other cities were treated as suburban to that nucleus. There are two exceptions to this rule. Both the twin cities of Minneapolis–Saint Paul and San Francisco–Oakland, because of their size and the availability of data for all four cities, were taken as single centers by summing the data for each respective pair.
2. Blalock (1971) provides a convincing argument that, whenever the direction of causation between exogenous (independent) variables is ambiguous, we cannot reliably estimate indirect effects through those variables. This being the case with our models, we shall restrict our analysis to the direct effects of each independent variable on the dependent variables.

FINDINGS

Does the size of the suburban population have any bearing on service functions performed in the central city? Table 1 presents the path coefficients between each of the four general categories of central city service functions and the population sizes of the central city and suburban areas.[4]

Looking first at the direct effects of central city size on the central city service functions, we observe that only central city wholesale trade in 1958 exhibits a significant positive relationship with central city size. Business and repair services along with public services show little relationship with central city size. Also interesting is that, when suburban population size is held constant, a strong inverse relationship emerges between retail trade per capita in the central city and the population size of the central city.[5]

TABLE 2

Path Coefficients between Changes in Central City Service Functions and Changes in the Population Size of Central Cities and Suburban Areas, 1950-60 and 1960-70

	Central City Service Functions			
Population Change	*Retail Trade (1948-58)*	*Wholesale Trade (1948-58)*	*Business and Repair Services (1948-58)*	*Public Services (1950-60)*
1950-60:				
Central city	.54**	.41**	—.15	—.18
Suburban area	.47**	.71**	.69**	.65**
Population Change	*Retail Trade (1958-67)*	*Wholesale Trade (1958-67)*	*Business and Repair Services (1958-67)*	*Public Services (1960-69)*
1960-70:				
Central city	.74**	.27**	.22*	.11
Suburban area	.31**	.62**	.51**	.45**

* Significant at .01.
** Significant at .001.

3. In simple recursive models such as ours, the path coefficients are equivalent to standarized partial regression coefficients.
4. For cross-sectional analysis, the logs of the size of the central city and suburban populations were regressed on the per capita service measures, since a preliminary plotting of these per capita measures with central city size and suburban size showed a slight bending effect at the upper ends of the distributions.
5. This, of course, does not mean that central city retail trade declines with increases in central city size but rather that, controlling for the effects of the suburban population, increments in central city population are associated with less than proportionate increases in central city retail trade.

On the other hand, all four categories of central city service functions exhibit highly significant positive relationships with the size of the suburban population, both in 1960 and 1970. Moreover, with the exception of wholesale trade at the former point in time, suburban population has a larger direct (positive) effect on every category of central city service functions than does central city population.

Having noticed the close association cross-sectionally between central city service functions and population size of the suburban areas, the question now becomes, To what extent are changes over time in these service functions influenced by, or related to, changes in the suburban population size? This is determined in Table 2 which shows the path coefficients between changes in the four categories of central city service functions and changes in the population of central cities and suburban areas during the two most recent decades.

The longitudinal results indicate that changes in the size of the suburban population have had highly significant direct effects on all four categories of central city service functions. The proportion of variance in the four service categories for which changes in the suburban population were directly responsible ranged from 22% to 51% during the 1950-60 decade, and from 10% to 38% during the 1960-70 decade.[6] It is again important to observe that changes in the suburban population exerted much larger direct effects on every central city service category, with the exception of retail trade, than did changes in the central city population.

The path coefficients presented in Tables 1 and 2 clearly indicate that the suburban population has a strong influence on central city service functions. Perhaps the most notable finding is that virtually no relationship exists between noneducational public services provided by the central city government and its population, either cross-sectionally or longitudinally, while positive and strong relationships exist both cross-sectionally and longitudinally between increases in the suburban population and central city public services. The obvious and important inference suggested by this finding is that the suburban population has at least as great an impact on public service provided by central city governments as does the central city population itself. The remainder of the study will examine this issue in more detail.

THE PUBLIC SERVICE ISSUE

One of the most serious problems facing our central cities in recent years has been the growing service-resource gap that has developed from a disproportional increase in public services provided by central city governments over that of their resources. Either explicit or implicit in most discussions addressing this problem has been the suggestion that the major force behind the increasing demand for public services is the changing composition of the central city population. Much less emphasis has been given to the increased demand for public services in the central cities created by the rapidly expanding suburban populations.

It would be difficult to deny that the changing composition of the central city population has increased the need for certain municipal services, such as public welfare and housing. However, we should not overlook the fact that increases in suburban populations have created a large demand for many other central city services. For example, the suburban population makes regular use of central city streets, parks, zoos, museums, and other public facilities; its routine presence in the central city increases problems of the sanitation department and contributes to the costs of fire protection; the daily movement in and out of the central city of the large commuting population requires services that constitute a large portion of the operating budget of both the police and highway departments (Hawley 1957, p. 773). These are only some of the costs experienced by central city governments as a result of services they provide to their suburban neighbors. Just what has been the relationship between growth of suburban populations and services commonly provided by the central city?

In Table 3 are presented the path coefficients between the sizes of the central city and suburban populations and the annual per capita expenditures for six common central city service functions.[7] We see that the suburban population exhibits a higher positive relationship with every central city service function than does the central city population. Moreover, while the direct effects of the suburban population are significant at the .001 level on every central city service function, except highway services in 1960, not a single significant positive relationship exists between the central city population and the central city service functions when suburban population size is held constant. These data provide support, in the cross-sectional case,

6. The squared path coefficients measure the proportion of variance (or change) in the dependent variable for which the determining variable is directly responsible (Land 1969, p. 10). The zero-order correlations between changes in central city and suburban populations were .11 during the 1950-60 time period and —.27 during the 1960-70 time period.

7. Police service includes police patrols and communications, crime-prevention activities, detention and custody of persons awaiting trial, traffic safety, vehicular inspection, and the like. Fire-prevention services include inspection for fire hazards, maintaining fire-fighting facilities such as fire hydrants, and other fire prevention activities. Highway services include maintenance of streets, highways, and structures necessary for their use, snow and ice removal, toll highways and bridge facilities, and ferries. Sanitation services include street cleaning, collection and disposal of garbage and other waste, sanitary engineering, smoke regulation, and other health activities. Recreation facilities and services include museums and art galleries, playgrounds, play fields, stadiums, swimming pools and bathing beaches, municipal parks, auditoriums, auto camps, recreation piers, and boat harbors. General control includes central staff services and agencies concerned with personnel administration, law, recording, planning and zoning, municipal officials, agencies concerned with tax assessment and collection, accounting, auditing, budgeting, purchasing, and other central finance activities (source: U.S. Bureau of the Census 1970, pp. 64-67).

to the contention that increases in the size of the suburban population are a major contributor to increased expenditures for common service functions performed by central city governments.

The next step is to examine the direct effects of changes (over time) in the size of the central city and suburban populations on changes in expenditures (in constant dollars) for the six central city services. Table 4 lists the path coefficients between changes in each service function and changes in the central city and suburban populations during the 1950-60 interval and during the 1960-70 interval.

We observe that the impact of the suburban population on the six central city service functions is even larger longitudinally than was true cross-sectionally. The percentage of change in these service functions for which changes in the suburban population are directly responsible range from 26% to 64% during the 1950-60 interval and from 26% to 38% during the 1960-70 interval. Only during the 1960-70 interval do we find changes in

TABLE 3
Path Coefficients between per Capita Expenditures for Common Central City Service Functions and Population Size (Log) of Central Cities and Suburbs, 1960 and 1970

Central City Service Function	Population, 1960		Population, 1970[a]	
	Central City	Suburb	Central City	Suburb
Police	.15	.56**	.09	.61**
Fire	−.22*	.43**	−.35**	.57**
Highway	−.19*	.11	−.28**	.25**
Sanitation	−.01	.31**	−.15	.56**
Recreation	−.12	.24**	−.03	.37**
General control	.06	.33**	−.02	.41**

[a] Central city service data for 1969.
* Significant at .01.
** Significant at .001.

TABLE 4
Path Coefficients between Change in Expenditures (in Constant Dollars) for Common Central City Service Functions and Change in the Central City and Suburban Populations, 1950-60 and 1960-70

Central City Service Function	1950-60		1960-70[a]	
	Central City	Suburb	Central City	Suburb
Police	−.12	.74**	.13	.56**
Fire	.01	.79**	.17	.54**
Highway	−.13	.61**	.17	.51**
Sanitation	−.03	.72**	.18	.62**
Recreation	−.10	.80**	.23*	.54**
General control	−.10	.51**	.20*	.58**

[a] Central city service function change 1960-69.
* Significant at .01.
** Significant at .001.

central city population size exhibiting consistent positive relationships wit changes in the city services. Even during this time period, however, the direc effects of changes in the suburban population were much larger than th effects of changes in the central city population.[8]

OTHER CAUSAL FACTORS

Having demonstrated through both static and dynamic analyses that th growth of common central city service functions is strongly related t population increases in suburban areas, we now test for the possibility tha some third variable or combination of variables are responsible for the stron relationships between suburban populations and central city public services First, we must account for the effects of the age of the city. We know tha suburban populations in the older SMSAs are generally larger than those i younger SMSAs. It has also been found that "older cities bearing the stamp of obsolescence, high density, high industrialization, and aging inhabitants generate higher expenses than their size alone might have led one to suspect" (Vernon 1960, p. 172).

Another important variable is personal income of central city residents Cities whose residents have higher personal incomes are usually able to pro vide more and better quality services than those whose population has lowe personal incomes. It has also been long known that city-to-suburb migratio is closely related to personal income of central city residents. Residents witl higher incomes have a much larger choice of residential locations than do lo income residents who are often economically and socially confined to th inner city.

Finally, we must control for the racial composition of the central city One may suspect that certain municipal expenditures are either directly o

TABLE 5
Zero-Order Correlation Coefficients between Selected Characteristics of Central Cities and per Capita Operating Expenditures of Central City Governments and Size of Suburban Population, 1960

	Central City Characteristic		
	Age	Per Capita Income	Non-white (%)
Per capita operating expenditures in central city (all services)	.43*	.14	.17
Size of suburban population (log)	.65*	.16	.28*

* Significant at .001.

8. Cross-sectional and longitudinal analysis was also carried out, controlling for suburba population annexed by central cities between 1950 and 1960. With this control instituted, th direct effects of the suburban population on all service functions remained positive an strong cross-sectionally and actually increased in the longitudinal analysis.

indirectly related to the racial composition of the central city population. At the same time, the "flight to the suburbs" has been greatest in those cities that have experienced the largest influx of nonwhite migrants in the past 25 years.

To discover if each of the above three variables were positively related to both per capita operating expenditures for services provided by central city governments and population size of suburban areas, zero-order correlations were computed.[9] As may be observed in Table 5, all relationships are positive. Per capita operating expenditures for central city services were found to be positively related at the .001 level of significance to the age of the central city, but positively related at only the .05 level of significance to per capita income of central city residents and percentage nonwhite in the central city population. Similarly, the size of the suburban population was found to be positively related at the .001 level of significance to both the age of the central city and the percentage nonwhite in the central city population, but less so ($p = .05$) to per capita income of central city residents.

These results raise an important question: what is the relationship between suburban population size and per capita operating expenditures for central city public services when we control, not only for central city size, but also for age of the central city, per capita income of the central city residents, and the percentage nonwhite in the central city population? In addition, we might ask: what is the relationship between per capita expenditures for central city services and each of the three central city variables (age, income, percentage nonwhite) when the remaining two central city variables, as well as central city size and suburban size are held constant? Table 6 answers these questions.

The crucial finding in Table 6 is that the impact of the suburban population remains strong and in the hypothesized direction when controls are introduced for central city size, age, per capita income, and percentage nonwhite. With these variables held constant, the direct effects of the size of the suburban population on the central city service functions are all positive, with significance at the .001 level for every service function except highway services.

Central city size exhibits a negative relationship to all service functions except sanitation services when the other variables are held constant. The significant negative relationships found between central city size and a number of the services indicate that economies of scale may operate in the provision of these services.

Examining the direct effects of central city age, per capita income, and

9. The correlation and path coefficients presented in Tables 5, 6, and 7 apply to 157 SMSAs in 1960 for which complete data were available on central city age (operationalized as the number of decades since the central city first attained a population size of 50,000 or more), per capita income of central city residents, percentage nonwhite in the central city, and the number of suburban residents who commute to work in the central city.

percentage nonwhite on the service functions, we observe that age has a substantial direct effect on the total operating expenditures and on expenditures for fire and highway services, as well as on general control. Also as expected, personal income of the central city residents exerts a positive effect on all services, with significant direct effects on fire and highway services. Percentage nonwhite exerts significant direct effects on sanitation services and general control but, in contrast to the effects of age and personal income, exhibits essentially no relationship with fire and highway services.

In sum, Table 6 indicates that size of the suburban population, rather than size or composition of the central city population, is the most important determinant of central city expenditures for public services. It may also be inferred from Table 6 that, *ceteris paribus,* the overall per capita operating expenditures for central city services increase with the age of the central city and decline with increases in its size.

THE COMMUTING POPULATION

In an effort to refine the above analysis and determine the impact on central city public services of suburban residents who utilize central city

TABLE 6
Path Coefficients between per Capita Expenditures for Central City Services and Selected Variables, 1960

	Central City Population (log)	*Suburban Population (log)*	*Age*	*Per Capita Income*	*Non-white (%)*
All services	—.24**	.38**	.34**	.10	.09
Police	—.04	.60**	.02	.15	.16
Fire	—.51**	.45**	.30**	.20*	—.03
Highway	—.29**	.08	.19*	.19*	—.04
Sanitation	.05	.28**	—.19*	.14	.27**
Recreation	—.19*	.27**	.04	.10	.12
General control	—.22*	.33**	.23*	.18	.22*

* Significant at .01.
** Significant at .001.

TABLE 7
Path Coefficients between per Capita Expenditures for Central City Services and Selected Variables, 1960

	Central City Population (log)	*Commuters (log)*	*Age*	*Per Capita Income*	*Non-white (%)*
All Services	—.23*	.35**	.35**	.11	.08
Police	—.02	.52**	.06	.17	.16
Fire	—.52**	.44**	.32**	.21*	—.04
Highway	—.40**	.28**	.14	.18	—.08
Sanitation	.02	.31**	—.18	.14	.26**
Recreation	—.18	.24**	.03	.11	.12
General control	—.21*	.28**	.25**	.20*	.22*

* Significant at .01.
** Significant at .001.

services daily, data were obtained from the journey to work reports of the 1960 Census of Population on the total number of people in each SMSA who reside in the suburbs and commute to work in the central city. Path coefficients were again computed, substituting the number of commuters for suburban population size in the least-squares equations. The results are presented in Table 7.

We observe that the number of suburbanites who commute to work in the central city has a direct impact at the .001 level of significance on the total per capita operating expenditures for central city services as well as on per capita expenditures for each individual central city service. Recalling that highway services was the only central city function to which size of the suburban population was not significantly related (Table 6), it is noteworthy that, when number of commuters is used as the independent variable, a highly significant positive relationship emerges.

The fact that the overall results in Table 7 are so similar to those in Table 6 indicates that the number of suburban residents who commute to work in the central city corresponds closely with the size of the suburban population. When the zero-order correlation between size of the suburban population and the number of suburban residents who commute to the central city was computed, it was found to be .95. Regression analysis showed the unstandardized slope between suburban size and number of commuters to be .105. In other words, an almost perfect linear relationship exists between suburban population size and the number of suburban residents who commute daily to the central city, with each increase of 1,000 suburban residents leading to an additional 105 commuters. Furthermore, the ratio of suburban residents who work in the central city to the central city resident population increases with the size of the suburban population. A correlation coefficient of .46 exists between suburban population size (log) and the ratio of commuters to central city residents. Thus, as the size of the suburban population increases, not only do larger numbers of suburban residents daily utilize central city public services, but, more important, the proportion of suburban residents relative to central city residents who utilize city services also increases. On the average, there are 132 commuters using central city services per 1,000 central city residents. These findings, along with the results provided in Tables 3 through 7, offer empirical support to the argument that the rapid growth of suburban populations has contributed greatly to the increased demand and, hence, increased expenditures for common central city public services.

SUMMARY AND IMPLICATIONS

This study examined the relationship between suburban population growth and service functions performed in central cities of 168 SMSAs. While most sociologists acknowledge that the suburban population influences the service structure of central cities, the degree of that influence has often been underestimated. Both cross-sectional and longitudinal analysis demonstrate that the suburban population has a large impact on central city retail trade, wholesale trade, business and repair services, and public services provided by central city governments. More detailed examination of the public sector shows that the suburban population in general, and the commuting population, in particular, exerts strong effects on police, fire, highway, sanitation, recreation, and general administrative functions performed in the central cities. The impact of the suburban population remains strong when controls are introduced for central city size and age, annexation, per capita income of central city residents, and percentage of the central city population that is nonwhite.

What implications do these results have for the present and future planning of the metropolitan community? First, the findings indicate that central city officials and planners should be particularly attentive to trends in the population growth of their outlying areas when projecting future demands for central city services. As long as area specialization continues to increase within the metropolitan community, we can expect the impact of the suburban population on central city facilities and services to grow.

A second implication is that the suburban population, by its daily use of central city facilities, substantially raises the costs of municipal services. While suburban residents do partially reimburse central cities in some SMSAs through employment and sales taxes, it is not likely that these "user charges" generate sufficient revenue to cover the additional costs.[10] A strong case can therefore be made for consolidating the politically autonomous suburban units with the central city in the form of a metropolitan-wide government. With a single jurisdiction controlling the services and resources, not only would the tax load for the provision of municipal services be spread in a more equitable fashion throughout the metropolitan area, but economies of scale might also be realized. Heavy resistance from suburban populations to political reorganization, however, makes the outlook for consolidation in the near future quite pessimistic. For the time being, then, the only recourse

10. A recent study (Neenan 1970) of benefit and revenue flows between Detroit and six of its suburban municipalities, representing both residential and industrial suburbs, shows that the suburban communities enjoy a considerable net gain from the public sector of Detroit. Neenan's analysis indicates that the net subsidy from Detroit ranges from $1.73 per capita for the low-income industrial suburb (Highland Park) to $12.58 per capita for the high-income residential and commercial suburb (Birmingham). Although not specifically analyzed in the present study it should also be noted that many central cities indirectly subsidize their suburban areas by having to pay an unfair share of the metropolitan area's welfare services. Through zoning restrictions and discriminatory practices, the suburban populations have been able to insure that most of the low-income, poorly educated, and chronically unemployed people in the metropolitan area are confined in the central cities. Suburban residents are therefore able to avoid the costs of public housing, public health, and other welfare services, which often impose a heavy burden on the operating budget of central cities.

open to the central city is increased financial assistance for the provision of municipal services. Perhaps suburban resistance to consolidation will only recede when the circuitous flow of taxes from suburb to Washington to central city increases to an extent that the service-resource gap begins to favor the central city.

REFERENCES

Blalock, Hubert M. 1971. "Path Analysis Revisited: The Decomposition of Unstandardized Coefficients." Unpublished research note. Department of Sociology, University of Washington, Seattle.

Hawley, Amos H. 1957. "Metropolitan Population and Municipal Government Expenditures in Central Cities." In *Cities and Society,* edited by Paul K. Hatt and Albert J. Reiss. New York: Free Press.

Land, Kenneth. 1969. "Principles of Path Analysis." In *Sociological Methodology 1969,* edited by Edgar Borgatta. San Francisco: Jossey-Bass.

Neenan, William. 1970. "The Suburban-Central City Exploitation Thesis: One City's Tale." *National Tax Journal* 23 (June): 117-39.

U.S. Bureau of The Census. 1970. *City Government Finances in 1968-69.* Washington, D.C.: Government Printing Office.

Vernon, Raymond. 1960. *Metropolis 1985.* Cambridge, Mass.: Harvard University Press.

Personal Journal

B. THE USE OF CENSUS DATA IN SECONDARY ANALYSIS: THE CONTEXT OF ECOLOGICAL DISCOVERY

THE CENSUS AS A SOURCE OF SECONDARY DATA

There is no larger or perhaps more important set of secondary data available for social research than censuses tabulated by the U.S. Bureau of the Census.[1] While the largest and most diverse single source of secondary data is the decennial (ten-year) Census of Population and Housing, the Bureau also conducts quinquennial (five-year) Censuses of Business, Manufacturing, Agriculture, Mineral Industries, Construction, Transportation, and Governments. These censuses provide a broad base of social science information ranging from the number of female-headed households in Albuquerque, New Mexico (Census of Population and Housing) to expenditures for police protection in Scranton, Pa. (Census of Governments). Data may be obtained to analyze educational attainment, ethnic and racial segregation, poverty, labor force participation, urban expansion, discrimination in employment and housing, commuting patterns, public services, marriage and divorce, fertility, migration, social mobility, and feminine careerism, to mention a few. So extensive are the data provided in the various censuses that opportunities for secondary analysis seem to be limited only by the imagination of the researcher.

What makes census data particularly useful for secondary analysis is that most variables are available not only cross-sectionally (i.e., at one common point in time) but also on a longitudinal basis (i.e., at a number of different points in time). This enables researchers to measure particular social relationships and examine their trends historically. For example, one may compare socioeconomic characteristics of different ethnic or racial groups over ten year intervals to determine if gaps in their socioeconomic status are widening or declining; or one may examine the suburbanization of industrial activity since World War II and analyze its impact on minority unemployment in the central cities. Most census data are now available for at least three points in time, with some comparable data extending as far back as 1790, the year of the first U.S. Census of Population.

Virtually all U.S. census data are tabulated on a geographic summary basis. This means that the data are aggregated (summed) for particular geographic areas, such as cities or counties. Population and housing data are tabulated for numerous and various geographic areas, ranging from separate residential blocks to the entire nation. The most commonly utilized geographic summary areas are blocks, census tracts, towns, cities, counties, metropolitan areas, regions, and states.

The fact that census data have generally been tabulated on a geographic summary basis has often precluded using individuals or households as units of analysis. In 1964, however, the Bureau of the Census made available on computer tape and punch cards separate records (information files) on 180,000 individuals drawn from the 1960 Census of Population and Housing. This sample included detailed information on one out of every 1,000 individuals in the United States. To ensure confidentiality, the names, addresses and other information that might possibly identify a particular individual were not included. In 1972 data were released from the 1970 Census of Population and Housing on approximately 2 million individuals (a 1 percent sample of the entire population). Again to avoid disclosure, all personal identifying information was removed from the records. These massive

Source: Prepared especially for this volume.

1. Secondary data may be defined simply as information on people, places, or topics drawn from sources initially assembled by somebody else. Analysis of these data, which is referred to as secondary analysis, usually focuses on issues other than those for which the data were originally assembled. For an excellent discussion of the principles, procedures, and potentialities of secondary analysis, the student may wish to refer to Herbert Hyman's recent book, *Secondary Analysis of Sample Surveys* (New York: Wiley, 1972).

samples of detailed social and economic data on individuals and households provide an extremely rich source of secondary data that is just beginning to be tapped by social researchers.

A fundamental advantage of using census data for social research is that the investigator is spared the time and cost involved in survey design, collection, and tabulation of data. The Bureau of the Census carries out these initial steps, enabling the researcher to proceed almost directly to the analysis stage. Furthermore, sampling and measurement errors in U.S. censuses are among the lowest of all sources of survey data. Census data are tabulated by an experienced staff of Bureau personnel who strive to achieve a high degree of completeness and accuracy for all data reported. Nevertheless, human and mechanical errors do occur which reduce the accuracy of some reported data. In the words of the Bureau of the Census:

> Such errors include failure to obtain required information from respondents, obtaining inconsistent information, recording information in the wrong place or incorrectly, and otherwise producing inconsistencies between entries or interrelated items on the field documents. Sampling biases occur because some of the enumerators fail to follow the sampling instructions. Clerical coding and editing errors occur, and errors occur in electronic processing. . . .[2]

To mitigate these errors, Bureau of the Census personnel review enumerators' work, verify the manual coding and editing, check the tabulated figures, and utilize statistical techniques such as ratio estimation of sample data to control totals in the complete count.[3] Through such efforts, errors in the printed reports are usually kept at an acceptably low level so that secondary analysis of these data will not generally yield misleading results.

Perhaps the major shortcoming of census data for social research is that many variables of interest to sociologists are not reported in the censuses. For example, no information is collected about sentiments, attitudes, values, or beliefs. Researchers interested in social-psychological studies must therefore look elsewhere for secondary sources of data or conduct their own field studies or experiments.

On the other hand, demographers and human ecologists find the information provided in the censuses most appropriate for their research needs. Their studies rely on broad bases of data on social, economic, and demographic characteristics of population groups and geographic areas. In fact the nature, scope, and completeness of data utilized in most demographic and ecological studies are such that they could never be reproduced by the individual researcher.

Census data are particularly well suited for the ecological approach in sociology. The basic unit of analysis in this approach is a geographically

2. U.S. Bureau of the Census, *U.S. Census of Population: 1960, United States Summary, Characteristics of the Population,* Vol. 1, pt. 1, p. lxxxv. (Washington, D.C.: Government Printing Office, 1964.)
3. Ibid.

delimited population. The population being studied may include those living in given census tracts, cities, counties, metropolitan areas, states, regions, nations or any other defined subarea. Like demography, human ecology is concerned with the size, composition, and distribution of particular populations. Human ecologists, however, focus primarily on environmental, structural, and technological factors affecting given populations and the consequences of population size, composition, and distribution for social organization. My analysis of the impact of suburban population growth on central city service functions is an example of an ecological study utilizing census data to examine one organizational consequence of the size, composition, and distribution of population in metropolitan America. Let me briefly chronicle the evolution of ideas and events in my research experience to illustrate how census data can be used for social research.

THEORETICAL BASIS AND INITIAL DESIGN

My interest in the influence of suburban population growth on central city service functions grew out of a seminar on human ecology that I took in 1969 under Amos Hawley at the University of North Carolina. The focus of the seminar was on theoretical issues in human ecology, though we also spent a good deal of time discussing methodological techniques for empirically grounding ecological theory. The issue that particularly attracted my interest is the theory of ecological expansion. In short, ecological expansion is a twofold process involving, first, a movement of people outward from a center of settlement without their losing contact with that center; and, second, a development of service functions in the center to sustain activities throughout the expanded system. The expansion process produces a diffuse community having two loci: an inner locus, or city, in which service functions are concentrated, and an outwardly advancing periphery, or suburb, which complements the city by routinely drawing on its services.

For my seminar project, I decided to examine the theory of ecological expansion empirically. My first step in formulating a research strategy was to select appropriate units of analysis. Two criteria seemed important to me. First, I felt that the units of analysis should be conceptually linked to the territorial system characterized by the theory and, second, that relevant data on these units should be available to test empirical propositions derived from the theory. Standard Metropolitan Statistical Areas (SMSAs) as established by the Bureau of the Census clearly satisfied both criteria.

The Bureau of the Census defines an SMSA as an economically and socially integrated territorial system with a large population nucleus, i.e., a central city of at least 50,000 inhabitants. The boundaries of SMSAs are delimited by the county containing the central city and all contiguous (adjoining) counties that are urban in character and economically and socially integrated with the county containing the central city. The inner

locus of SMSAs is geographically delimited by the central city and the outer locus by the suburban ring, which is that area outside the central city but within the SMSA boundaries. Since the Bureau of the Census provides comparable social, economic, and demographic data for both the central city and the SMSA, it is usually possible to derive data on the suburban ring by simply subtracting the appropriate central city data from the SMSA data.

Having chosen SMSAs as my units of analysis, I directed my efforts towards examining the central city as a service center for the suburban population. In so doing I hoped to show that metropolitan growth is a special case of ecological expansion. Two rather closely related hypotheses were derived from the theory of ecological expansion. The first followed from the proposition that population growth in the peripheral areas of a territorial system will be matched with an increase in service functions in the nucleus to sustain the enlarged system. The research hypothesis I derived from this proposition is that service functions performed in central cities will increase at a disproportionate rate as the size of the suburban population becomes larger.

My second research hypothesis followed from the proposition that in an expanding system, people move out from the center of settlement without losing contact with that center. If metropolitan growth is a special case of metropolitan expansion, then suburban population growth should lead to larger numbers of people who reside outside the central city but maintain direct contact with the central city through their daily use of its facilities and services. In turn the daily use of the central city by larger numbers of suburbanites creates an increased demand for public services in the city. I therefore hypothesized that a major cause of increased expenditures for central city services such as police, fire, highway, sanitation, and recreation is the expanding suburban population which makes routine use of these services while in the central city.

DATA AND ANALYSIS

My next step was to gather data to test my two research hypotheses. A perusal of recent censuses showed that data were available on the size, composition, and distribution of populations in central cities and SMSAs from Censuses of Population, that data on central city and SMSA retail, wholesale, and business service activity were available from the Censuses of Business, and that detailed data on public services provided in the central cities were available from the Censuses of Governments and annual Compendia of City Government Finances. Combining secondary data from these sources over different time periods enabled me to empirically test my hypotheses both cross-sectionally and longitudinally.

Because my time and budget were limited, I restricted my initial sample to those SMSAs that had a central city population of at least 100,000

($N = 91$). Still, with a sample size of 91 SMSAs and almost 20 variables for each unit, analysis by hand or desk calculator seemed out of the question; the computer became a necessity. This meant, however, that I had to keypunch the data from the various census volumes onto machine-readable cards (i.e., IBM punch cards). I can recall quite well the many tedious evenings I spent at the computer center in Chapel Hill keypunching and verifying the data. I must admit that at times I questioned the value of such efforts, particularly since I was not certain that anything meaningful would result in the final analysis. After about six weeks of coding, keypunching, and verifying, my data were ready to be analyzed.

In the analysis stage, I relied on a very useful canned computer program called SPSS (Statistical Package for the Social Sciences). This program enabled me to carry out basic cross-tabular and correlation analyses of all the dependent and independent variables I had coded. One of the most satisfying phases of the research project was discovering that the results of the analysis strongly supported both my research hypotheses. I wrote my findings up as a term paper, which I presented the following summer at the annual meetings of the American Sociological Association under the title "The Impact of Suburban Population: An Analysis of Ecological Correlates."

Encouraged by the results from my analysis of a sample of SMSAs, I presented a research proposal for financial support to expand the study to the Department of Sociology at the University of North Carolina. In the summer of 1970 I was awarded $600, which I used to hire a research assistant to code, keypunch, and verify additional census data. The data collection was extended to include all SMSAs as of 1950 ($N = 168$), recently released 1970 census data, and other data which I felt were essential for statistical control.

There were two methodological problems in my original analysis that I hoped to solve, in part, with the expanded data base. First, I had been using sales, receipts, and expenditure data (measured in dollars) as indicators of service activity in the central cities. Since the value of the dollar substantially changed between 1950 and 1970, it was essential to transform the sales, receipts, and expenditures into constant dollars. In transforming these data, I utilized the consumer price index to deflate retail sales and business service receipts, the wholesale price index to deflate wholesale sales, and the purchasing power of the dollar to deflate expenditures for public services.

The second methodological problem that became apparent was that although I had previously demonstrated a strong correlation between suburban population growth and central city services, the two factors were not necessarily functionally related. It is entirely possible for two variables to be highly correlated yet functionally (or causally) unrelated. When this occurs we have what statisticians term a "spurious" relationship. A spurious relationship between two variables emerges whenever some third variable or set of variables causes two functionally unrelated variables to co-vary in a

systematic manner. A commonly cited spurious relationship is that between ice cream sales and crime rates. A strong positive relationship exists between these two factors because both vary systematically with weather conditions, rather than one factor being the cause of the other.

To guard against possible spurious relationships I found it necessary to control for other variables that may affect both the size of the suburban population and central city service functions. My review of the literature and preliminary correlation analysis suggested three variables that may result in a spurious relationship between the size of the suburban population and expenditures for central city services. As discussed in the article, these variables are the age of the central city (measured in terms of the number of decades passed since the city attained a population of 50,000 or more), the per capita income of central city residents, and the percent of nonwhites in the central city. If, after controlling for these variables, a strong relationship remains between the size of the suburban population and central city service functions, then it is quite likely that a "real" rather than a "spurious" relationship has been discovered.

How, though, does one determine the strength of relationship between two variables when controlling for other theoretically relevant variables? The method that I chose was path analysis, which is no more than basic regression analysis including a number of theoretical assumptions about linearity, additivity, and causality.[4] A path coefficient is simply a standardized regression coefficient which is often referred to as a "Beta weight" in the research literature. It indicates how much variance in the dependent variable (e.g., public service expenditures) is produced by a standardized change in one of the independent variables (e.g., suburban population size) when all other independent variables under analysis are held constant. Through path (regression) analysis I was able to determine the direct effects of not only suburban population size but also the direct effects of the age of the central city, percent nonwhite in the central city, and the per capita income of central city residents.

My analysis of the direct effects of all these variables indicated that the suburban population has by far the most pervasive impact on expenditures for public services in central cities. Of course I was pleased that the results supported my contention that the size of the suburban population is at least as important a factor influencing central city service expenditures as is the population composition of the central city itself. However, the article's concluding remarks reflect my feeling that the results were even more important from a standpoint of public policy. While I still believe the findings imply that central city planners and officials should be more cognizant of population trends in their suburban areas when projecting future demands for central city services, I now feel that I speculated beyond the evidence of my data by implying that the suburban population may be exploiting the central city with respect to public services.

No doubt the results suggest that the suburban population increases the costs of public services for the central cities. But to accurately infer exploitation I would also have to demonstrate that these costs are greater than the revenue the suburban population generates for the central cities. The suburban population makes contributions to the central city resource base other than commuter taxes, sales taxes, or other "user charges." For example, suburbanites' employment in downtown office buildings indirectly improves the central city tax base by increasing property values. Moreover, many central city business establishments could not survive without the large number of suburbanites who purchase their goods and services. The whole issue of central city-suburban exploitation remains a problem yet to be definitely resolved in the research literature.

FURTHER REFLECTIONS ON THE STUDY

When a researcher completes a study, he often looks back and asks himself what he would do differently to improve the study. There are two changes that I feel would make my analysis more convincing and important from a policy standpoint. First of all, I would add more independent (control) variables to my path models. The one criticism frequently heard about path (regression) analyses such as mine is: "But would your results have been the same if you had controlled for *X*?" For example, including a measure of the quality of public services in the central cities would have improved the analysis. Perhaps public service expenditures are higher in central cities with larger suburbs, but their services are also of better quality. Here the possibility of spuriousness is again raised. Another important variable that I would now include in the analysis is federal and state financial aid to the central cities. Recent research suggests that expenditures for some city services are quite sensitive to the amount of extralocal subsidies the cities receive.

A second step that I would take to improve the study would be to conduct a cost-benefit analysis of the use of central city services and facilities by the suburban population. This would involve computing estimates of the economic contribution that suburbanites make to the resource base of the central city along with estimates of the marginal costs they create in making use of central city services. Regression analysis of data from the Censuses of Population, Business, and Governments is one method that could be used to estimate the costs and benefits that accrue to the central cities as a result of their use by the suburban population.

4. For a straightforward overview of path analysis see Kenneth Land, "Principles of Path Analysis" in Edgar F. Borgatta, ed., *Sociological Methodology 1969* (San Francisco: Jossey-Bass, 1969), pp. 3-37.