Volume Title: Annals of Economic and Social Measurement, Volume 1, number 2

Volume Author/Editor: Sanford V. Berg, editor

Volume Publisher: NBER

Volume URL: http://www.nber.org/books/aesm72-2

Publication Date: April 1972

Chapter Title: The Role of Data Availability in Intrametropolitan Workplace Location Studies

Chapter Author: Robert A. Leone

Chapter URL: http://www.nber.org/chapters/c9192

Chapter pages in book: (p. 171 - 182)

Annals of Economic and Social Measurement, 1/2, 1972

THE ROLE OF DATA AVAILABILITY IN INTRAMETROPOLITAN WORKPLACE LOCATION STUDIES*

BY ROBERT A. LEONE

Previous research relating to intrametropolitan location patterns has been generally deficient primarily due to the lack of an appropriate data base. This paper first identifies the most glaring deficiencies of the data sets commonly used for location analysis; second, it enumerates a set of criteria for a sound data base; and third, it investigates the potential of a new data base, the Dun & Bradstreet DUNS Market Identifier (DMI) file as one readily available and reasonably accurate data source which comes close to satisfying these criteria. To illustrate the usefulness and flexibility of the DMI file, an example of how the file was used in a recent study of land use patterns in the New York Metropolitan Area will be discussed.

DEFICIENCIES IN COMMONLY USED DATA SETS

Data sets commonly used to study location behavior (e.g., the Census of Manufactures and transportation system survey data) generally have three deficiencies: (1) they frequently lack the spatial detail necessary to identify the geographic direction of land use change within metropolitan areas; (2) they quite often fail to have sufficient establishment detail to permit the identification of the process of change and its contributing factors; and (3) survey data are frequently cross sectional, without linkages over time.

The lack of geographic detail has long plagued researchers who are well aware that economic phenomena rarely follow political jurisdictions. In the analysis of urban growth, for example, an inability to designate such commonly utilized geographic classifications such as a Central Business District (CBD) or a Central Industrial District often has been a serious handicap. Census tract information helps, of course, especially in residential location analysis, but rarely are tracts sufficiently homogeneous to permit the isolation of even the most important factors affecting land use patterns, not to mention the more subtle forces which shape urban development.

It is easy to illustrate the importance of geographic detail. The County of Manhattan, by necessity, is frequently taken to be synonymous with the Central Business District of New York City. For many purposes this assumption is harmless. Yet, it is extremely interesting to note that when refined geographic data are available, important differences manifest themselves. For example, manufacturing employment on the island of Manhattan grew by 9 percent between July, 1967 and August, 1969. However, the CBD (Manhattan south of Central Park) experienced an employment growth rate of 10 percent while Manhattan outside the CBD actually experienced an employment decline of over 29 percent.[1] The lack of geographic detail obscured this markedly different behavior.

* Received October 1971.

[1] The Dun & Bradstreet data permitted this distinction to be made. See Robert A. Leone, "The Location of Manufacturing Activity in the New York Metropolitan Area," unpublished Ph.D. dissertation at Yale University.

The lack of establishment detail has also hampered research. In the New York Regional Plan studies of the late 1950's, for example, Hoover and Vernon[2] were able to suggest a number of explanatory hypotheses of location behavior but almost always found conclusive tests of the hypotheses elusive usually because of insufficient establishment detail. Decentralization of economic activity is frequently identifiable from aggregate statistics, but whether or not decentralization is a characteristic common to all types of activities is extremely difficult to determine. For example, the tendency has been to equate decentralization in the aggregate with the lack of economic viability of the central city. More detailed analysis, however, suggests that much more subtle processes are actually occurring. Aggregated data sources reveal the result, but obscure the process of change. Both of these deficiencies can be corrected with micro data on individual location decision units. In the case of work place location studies, the establishment is the relevant decision unit. With micro data on individual establishments the researcher can identify any geographic area of his choosing. The greater the establishment detail the better the identification of the forces at work reshaping urban land uses.

The deficiency of cross section data, however, stems not from lack of establishment detail alone but also from the fact that the location pattern at any point in time is the cumulative result of decisions made over a long period of time. The long-lived character of many location decisions suggests that predictions of future location patterns based on the average behavior summarized in cross section data will always bias predictions in favor of the existing land use configurations while obscuring the marginal factors currently operative at the margin. This would not be a problem if there was any reason to believe that average behavior and behavior were similar; however, the well documented trend to decentralization clearly indicates that marginal behavior is markedly different from average behavior.

This deficiency can be rectified by utilizing cross section data over time. Aggregate cross section data over time, however, are inadequate. Although the analysis of net changes in aggregates is an improvement over single period cross section data alone, it does not permit the identification of the forces operating to create change. Net changes may be comparatively small, but the consequence of substantial, but largely counterbalancing, change. Only micro level data which permit the monitoring of specific establishments over time permit the desired disaggregation of net change into its component parts.

Criteria for a Good Data Set

The above observations suggest four criteria that a data set should satisfy if it is to be appropriate for the study of workplace location decisions. First, geographic detail is essential. Not only is physical location desirable, but numerical coding is desirable in order to facilitate data processing. Further, location codes should be mutually exclusive. The postal Zip Code represents one convenient and inexpensive codification scheme that satisfies these criteria, albeit at too great a level of aggregation for some purposes.

[2] Raymond Vernon and Edgar M. Hoover, *Anatomy of a Metropolis* (Anchor Books, New York, 1962).

Second, sufficient establishment detail to permit the identification of the process of change is needed. Ideally, quantitative measures of the following six location factors should be available by establishment:

(1) material assembly requirements;

(2) labor assembly needs by skill;

(3) land assembly needs with specification of any peculiar topographic requirements;

(4) the relative importance of face-to-face interaction with suppliers and customers;

(5) inter- and intracity transportation requirements; and

(6) service assembly needs.

Although the ideal data set would specify "needs" in the form of demand curves, a clearly acceptable second-best alternative would be to have actual levels of factor use.

More significant problems occur when certain needs are difficult to measure and require the use of proxies. For example, service assembly needs are probably the most difficult to measure quantitatively. The relative importance of face-to-face communication is also difficult to quantify.

One way to capture these forces might be to cross classify establishments by industry, size and area of primary functional responsibility. Industry classification would permit isolation of some of the technical factors affecting location. If combined with material coefficients, the linkages among industries could also be identified, thereby identifying some of the potential areas for agglomeration economies. In addition, agglomeration economies are frequently associated with establishment size which might be measured in terms of employment, sales, or net worth. A proxy for communications needs might be establishment function. For example, headquarters facilities would presumably be in greater need for the communications advantages of the central city than an operating facility of the same size in the same industry.

The third need is for cross section data over time. This would permit the detailed specification of change in the location pattern.

A fourth desirable attribute of any micro data set is that its use should not entail disclosure problems. Users of micro data are well aware of the obstacles to research that restrictions on disclosure can consitute.

The Dun & Bradstreet DMI File

A data set meeting many of the above requirements is available from Dun & Bradstreet, Inc. at moderate cost.[3] In particular, these data permit the monitoring of location changes over time at the micro level. What follows is a brief description of this data set, followed by an illustration of how rather unsophisticated data amplification techniques can augment an already rich data source. And finally, a simple quality check of the DMI file will be discussed.

[3] For pricing details contact Dun & Bradstreet, Inc., 99 Church Street, New York, New York 10007. To illustrate the costs of purchasing the data, individual establishment records are available on IBM 360 compatible magnetic tape at a rate of $300/1,000 establishments. All U.S. manufacturing establishments (approximately 400,000 in number) are available for $28,500. D & B's entire file of all 2,613,000 business establishments in the United States costs $132,500. These figures are, of course, unofficial.

The Duns Market Identifier (DMI) file represents an attempt by Dun & Bradstreet to compile a virtually exhaustive cross section sample of manufacturing establishments operating in the United States to facilitate identification of prospective customers by potential suppliers. The following items are reported by establishment:[4] (1) one primary and up to five secondary 4-digit SIC code classifications of products manufactured; (2) size information, including total employment for the establishment, total employment, net worth and sales for the firm of which the establishment is a part; (3) establishment activity information, including designation of headquarters, branch and subsidiary establishments, (4) geographic information, including the non-overlapping numerical codings of Zip Code, city, county and state; (5) year of formation of establishment, and lastly, (6) a "DUNS" (*D*uns *U*niversal *N*umbering *S*ystem) number, a seven digit integer identifier which is constant for any establishment through time. Because this number is never reissued it is relatively simple to trace the location history of any particular establishment.[5]

A problem with the DMI file. Unquestionably, a major limitation of the Dun & Bradstreet DMI file stems from the fact that historical files are available only since 1965. In order to create a time series of cross sections it is necessary to subscribe to the DMI file service over time. Updates are available on a quarterly basis and three quarterly updates are usually included in the initial purchase price.

Clearly, the lack of historical files prohibits analysis far into the past. The rapidity with which the workplace location pattern changes, however, goes a long way toward mitigating what would otherwise constitute a severe limitation on the usefulness of the file. In a recent two year period, for example, over 4,500 manufacturing establishments relocated in the New York Standard Metropolitan Statistical Area. This represented a shift in the geographic location of nearly 10 percent of the area's manufacturing jobs. During this same period, 3,000 manufacturing establishments ceased doing business in the New York area and an additional 4,000 started operations.

Experience clearly indicates that large files of information on marginal workplace location behavior can be generated in short periods of time. Comparable rates of change have been experienced in other cities besides New York.[6]

DMI file accuracy. In discussing the accuracy of a micro data set it is convenient to make the distinction between accuracy in the aggregate and micro level accuracy. As might be expected of any large micro data file, errors which are apparently minor when the statistics are aggregated can become rather substantial when dealing with individual components of that file. Estimates of GNP, for example, are frequently quite accurate, while the estimates of its component parts have rather high variances.

[4] Establishment definitions conform to Census criteria. Data on wholesalers and retailers are also available, but have not been used by this author.

[5] Certain individual pieces of data are omitted from some establishment records. For example, branch establishments have no date of formation figures.

[6] There are three separate studies of intrametropolitan work place location behavior which employ the DMI file. Robert A. Leone, "The Location of Manufacturing Activity in the New York Metropolitan Area," unpublished Ph.D. dissertation at Yale University; Gordon Saussy of the department of Business and Economics at the University of Louisiana at New Orleans is currently studying location behavior in New Orleans using the DMI file data; and Raymond Struyk and Franklin James "Four Cities," National Bureau of Economic Research, 1971 (mimeograph).

Given that the need for micro data in the analysis of workplace location patterns was justified on the grounds that marginal location decisions are more revealing than cross section averages or even net changes in aggregates, the value of the DMI file for location analysis is largely dependent on the accuracy of individual records.

Unfortunately, precisely because the DMI file is unique, it is extremely difficult to determine the accuracy of individual entries. Furthermore, errors in the data set can be either errors of commission or omission. Data for establishments can be incorrect or omitted from the survey altogether. It is especially difficult to determine the accuracy of individually cited statistics without replicating part of the survey. On the other hand, it is possible to get a feel for the number of establishments omitted from the survey. In the analysis of marginal location decisions this is the more important consideration.

The dominant importance of errors of omission is rather easily illustrated in Figure 1. To identify marginal location decision units, it is necessary to be able to identify the location of a particular establishment in two points in time. If, in both points of time, it has the same location, then it is clearly not one of the marginal location decision units. Conversely, if its location has changed over time, then it clearly is a marginal decision unit.

Problems arise when an establishment is identified in one but not in the other time period. For example, if the file is perfectly accurate and an establishment is

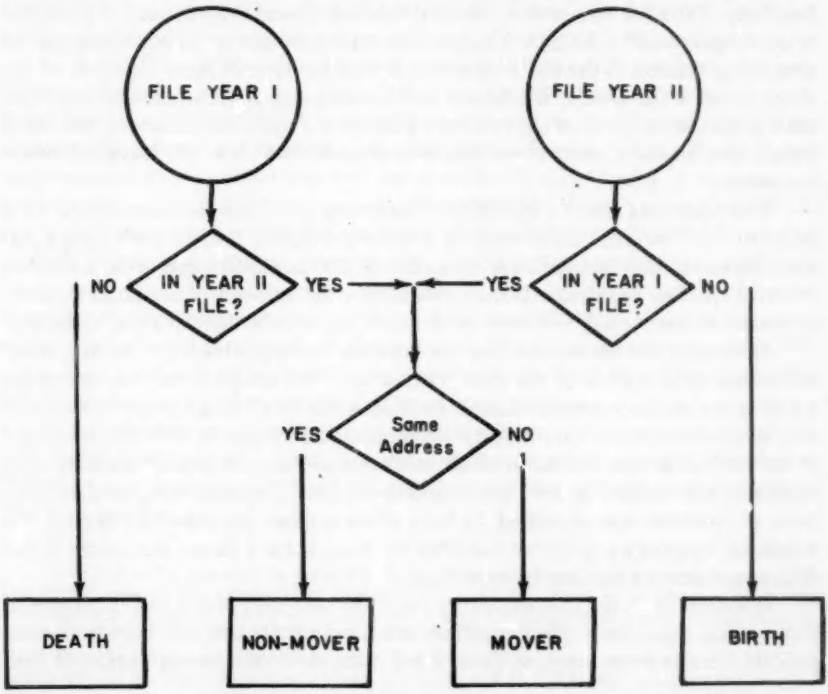

Figure 1 Creation of files of Establishments which recently made a location decision

TABLE 1

FINAL SUBFILE ESTABLISHMENT COUNTS

	Total	Both	Year(s) Observed: 1967 Only	Year(s) Observed: 1969 Only
Non-movers	31,827	25,734	1,439	4.654
Movers	4,501	2,992	872	637
Births	4,011			4,011
Deaths	2,970		2,970	
Total observations:			1967	1969
		Actual	34,007	38,028[2]
		Dummy	5,291	2,311
		Error		−26[1]
		Total	39,298	40,313

[1] There are 4,501 mover destinations, but due to a programming error only 4,475 actually appear in the 1969 figures. Total employment undercount in 1969 due to this error is 1,007.

[2] Original figure was 38,031. Three establishment records were lost during data correction.

located in the DMI file in year I but is not found in year II (deletion to coverage), the establishment either relocated outside the area under consideration or ceased to do business. On the other hand, if the data set is imperfect, it is possible that Dun & Bradstreet simply failed to survey that particular establishment in year II. Similarly, if the file was perfect, an establishment found only in year II (addition to coverage) would have to be either a new establishment or an immigrant to the area being studied. If the file is incorrect, it may be none of these. In either of the above cases, if the Dun & Bradstreet survey technique is accurate, any establishment found in only one of the two time periods is a marginal locator. If the file is innacurate, however, there is nothing to guarantee that these are marginal establishments.

The problems created by establishment omissions are far from minor. In a study of the New York Metropolitan Area in the period 1967 to 1969, over 8,300 establishments in a base of approximately 40,000 establishments were additions to coverage. The usefulness of the DMI file for the analysis of marginal location decisions depends to a great extent on the accuracy of these deletions and additions.

Obviously, the accuracy of the file can only be determined by referring to an additional data source. In the New York study[7] the simplest way to determine whether the numerous establishment omissions were legitimate or not was to scan area telephone directories. If an establishment deleted from the 1969 file was found in the 1969 telephone directory, an error of omission was identified. Similarly, if an establishment omitted in 1967 was found in the 1967 telephone directory, another error of omission was identified. In both of these cases the information from the telephone directory augmented the DMI file data. Table 1 shows the results of this data augmentation scheme for New York.

It should be noted that this procedure does not even direct itself to errors of commission nor does it affect establishments incorrectly omitted from both time periods. On the other hand, as Table 2 indicates, this rather unsophisticated data

[7] Leone, *op. cit.*

TABLE 2

EMPLOYMENT COUNT NEW YORK SMSA[1]

Area	*A* Census	*B* D& B[1]	*C* Percent Original[4]	*D* After[5]
SMSA	1,147,400	1,147,808[2]	90.0	100.0
Brooklyn	220,300	210,932	86.4	95.7
Manhattan	482,300	489,257	93.6	101.4
Queens	132,300	129,525	91.8	97.9
Richmond	7,700	7,806	90.3	101.3
Bronx	52,700	48,573	78.2	92.2
Suffolk	48,800	44,665	86.3	91.6
Nassau	116,000	103,435[3]	86.9	89.2[3]
Rockland	14,000	18,752	96.9	134.0
Westchester	73,300	71,863	90.0	98.1

[1] Employment must be estimated from the D & B data when only a range is indicated. In this case, the centroid value was used.

[2] Includes correction in Nassau county. See footnote 3.

[3] 23,000 undercount in transportation appears possible. With this correction, the percentage becomes 109.0.

[4] Ratio of original D & B statistics (not shown) to census figures in Column A.

[5] Ratio of corrected D & B figures to census figures: column B divided by column A.

amplification process resulted in a substantial improvement in the overall accuracy of the data set and presumably in the accuracy of the files of marginal locators.

Column D in Table 2 shows the relationship between the corrected aggregate statistics by county in the New York SMSA and the 1967 Census of Manufactures data. The total for the SMSA is embarrassingly similar. In all cases, the county figures were improved by using the telephone directory data. Further, the understatement or overstatement of the statistics in each county is rather easily understood in terms of the data imputation procedure. When a record was incorrectly omitted from the 1969 file, the employment observed in 1967 was imputed to 1969. Similarly, omissions from the 1967 file resulted in imputations of the 1969 data. Since Dun & Bradstreet is continually improving the coverage of its file, there are more imputations from 1969 to 1967 than the other way. Hence, areas which are growing the fastest, such as Rockland and Manhattan counties, are more likely to have 1967 overestimates and areas such as Brooklyn or Queens which are declining, are more likely to have underestimates.

It would have been possible to use a more sophisticated imputation technique. The improvement in the aggregates possible from a more complex imputation scheme was purposely sacrificed in order to keep imputations in the marginal files to a minimum.[8]

[8] This is a rather conservative approach. For example, employment increases of almost 22,000 jobs were associated with 4,500 moving establishments in the New York study. The conclusion that movement was strongly associated with growth was a highly conservative statement because many of the movers showed no employment change because of the imputation procedure. Had those establishments observed in only one time period grown at the same rate as those observed in two time periods then growth in moving establishments would have amounted to 32,000 jobs—substantially stronger confirmation of the conclusion that growth and movement were associated.

A macro quality check. To test the overall accuracy of the DMI file sample, statistics aggregated from the file were compared with figures published in the 1967 Census of Manufactures for the New York SMSA. There were no immediately apparent systematic differences. Comparisons of employment and establishment counts by industry, establishment size and geographic area are contained in the appendix.

This comparability with census statistics represents a basic source of confidence in the quality of the data. Further, because the data are compiled at the micro level and should be used in rather small subclassifications within limited geographical areas, obvious large errors and inconsistencies are easy to detect and relatively simple to take into account.

Findings in the New York Study

Although the purpose here is not to discuss the substantive results of the New York study, a brief summary of the findings will serve to indicate the potential of accurate micro data for workplace location analysis.

Table 3 perhaps best illustrates the value of the detailed geographic information. Broken down by geographic area and by marginal establishment category, employment changes in New York in the 1967–1969 time period were not, in fact, what might have been predicted on the basis of conventional wisdom. The New

TABLE 3

Employment Change by Marginal Establishment 1967–1969

Zone	Net Change	Births	Deaths	Change Local Movers	Movers		Non-movers
					In[1]	Out[2]	
Manhattan	+44,806	+24,995	−18,870	+8,499	+6,029	−10,626	+34,779
CBD	+48,878	+24,602	−17,890	+8,136	+5,851	−10,625	+38,804
Brooklyn	−1,624	+5,882	−8,083	+1,098	+5,777	−4,960	−1,338
Williamsburg	−1,202	+1,606	−1,216	+98	+1,538	−1,486	−1,742
Blytheborne	+556	+769	−26	+2	+6	−58	−137
Queens	−6,115	+3,713	−3,901	+961	+6,892	−3,078	−10,702
L.I. City	−7,174	+1,391	−1,876	+245	+3,298	−2,179	−8,053
Flushing	+729	+984	−1,042	+686	+3,262	−1,133	−2,028
Jamaica	+48	+1,267	−962	+57	+946	−507	−753
Bronx	−530	+1,451	−2,548	−285	+1,744	−1,498	+606
Richmond	−746	+821	−70	−77	+130	0	−1,550
Westchester	−9,932	+2,954	−2,688	+1,025	+1,440	−859	−11,804
Mt. Vernon	−6,489	+1,390	−1,559	+48	+104	−658	−5,814
Yonkers	−635	+1,300	−517	+986	+1,944	−612	−3,736
Nassau	+4,832	+3,185	−2,874	+3,163	+3,585	−3,116	+889
Valley Stream	−4,194	+1,253	−1,067	+159	+2,508	−1,653	−5,394
Hicksville	+14,970	+1,610	−1,606	+2,764	+1,894	−2,066	+12,374
Suffolk	+314	+1,528	−1,179	−138	+3,033	−1,042	−1,888
Babylon	+1,132	+1,418	−1,108	−138	+3,010	−1,042	−1,008
Rockland	+2,886	+617	−361	−62	+354	−366	+2,704
SMSA	+33,891	+45,146	−40,574	+14,184	+9,018[3]	−5,579[3]	+11,696

[1] Employment in 1969 figures.
[2] Employment in 1967 figures.
[3] Net of local moves.

York CBD, for example, was doing quite well; the relative importance of migration within the region—a much publicized central city "problem"—proved to be of much less consequence than some of the other components of change.

The benefits from establishment detail (not illustrated in the table) were as great as those from geographic detail. For example, when disaggregated by industry, the decentralization of manufacturing activity observed in the aggregate was not found to be a characteristic of all industries. In fact, many industries were rapidly centralizing, either by in-migration to the central city or by faster growth in the central city.

The DMI file data on establishment function was particularly revealing. Consistencies in location behavior were frequently more noticeable when establishments were classified by function without regard to industry than they were when classified by industry, but not by function.

Given the importance of understanding how land use patterns evolve in cities and the lack of research on the subject, it appears quite clear that the Dun & Bradstreet DMI file has much to contribute to workplace location analysis before the shortcomings of the data set became serious constraints.

Conclusion

The Dun & Bradstreet DMI file is not the answer to all data problems for intrametropolitan location studies. But it certainly represents a substantial improvement over commonly utilized alternatives. The file offers four obvious advantages to researchers attempting to disentangle location behavior:

(1) Although virtually a private census of manufactures, there are no disclosure problems. In fact, participants in the survey do so primarily because they want the information disclosed.

(2) The zip code level of geographic detail is sufficient for almost all uses. This item permits a level of geographic refinement previously unattainable. In addition, the presence of physical address permits those with the need for information on precise locations to make use of the sample, although at higher costs.

(3) Establishment detail is far greater than in any other readily available micro data set suitable for location analysis.

(4) The completeness of the file is remarkable, although the New York figures are potentially misleading. New York is the home city of Dun & Bradstreet and as such may have better coverage.

The potential the file offers for urban economic applications is obvious: detailed intraurban land use pattern changes can be monitored; regional shifts in activity can be analyzed at the micro level. The location of basic sector activity, so important to most urban growth models, can now be better understood.

There are, in addition, substantial possibilities for the DMI file outside the urban economics area. Firm behavior can be monitored on an establishment basis, for example. The DMI file can be merged with other existing micro, and even macro, data sets to create more nearly complete records of establishment behavior.[9]

[9] In one market organization study, data from the DMI file has been merged with data available from Compustat. See Michael Gort, *et al.*, "Firm Data and Industry Aggregates in the Analysis of Diversification and Integration," *Annals of Economic and Social Measurement*, January, 1972.

In sum, the Dun & Bradstreet DMI file represents a substantial improvement in both the quantity and quality of micro data available for workplace location analysis. Its ready availability and richness will undoubtedly whet researchers' appetites for more accurate and more comprehensive data sets, but in the meantime, there is much that can be done.

National Bureau of Economic Research
and Yale University

Appendix: Comparison of the DMI File with the Census of Manufactures

Perhaps the simplest way to check the quality of the figures found in the DMI file sample is to compare some of the aggregate statistics to a more frequently employed data base. The 1967 Census of Manufactures offers a convenient vehicle for this comparison. The census figures quoted below represent a substantial improvement over any alternative primarily because they were tabulated in 1967. Table 2 above shows the relationship of the aggregated county statistics in the two samples. On this basis the samples are quite comparable. No systematic differences are immediately apparent.

Table 3 above shows the distribution of employment by industry at the 2 digit level for the SMSA as calculated both by the census and the DMI file after verification. Both the raw employment figures and the percentage distribution are given. Some of the disagreement is apparently explained by the non-availability of certain employment data in 1967. For example, the undercount in Nassau county observed in Table 2 was due entirely to the failure to report employment for one large transportation equipment establishment. As a consequence, of course, the transportation equipment industry was substantially under-represented in industry distribution figures as well. Note, however, the improvement in both transportation equipment and chemicals in 1969.

Tables A-1 and A-2 also contain the figures on the distribution of establishment within the metropolitan area. Here there is more variability between the census figures and the Dun & Bradstreet figures. The similarity of the employment figures suggests that the variation in the establishment counts is due to different treatment of small, perhaps zero employment establishments, In Table A-2 the percentage distribution shown for the Dun & Bradstreet data is net of all establishments reporting zero employment to make the comparison with the census possible. Also shown are the numbers of zero employment establishments in the area.

A last crude comparison between these two data sources is shown in Table A-3. In that table the number of large establishments (establishments employing more than 240 or 250 persons) is shown by county. Unfortunately the columns are not strictly comparable because the census reports establishments employing more than 250 persons while our Dun & Bradstreet tabulations include establishments employing more than 240 persons. Clearly, the D & B total should always be greater than or equal to the census total and it in fact is except in Queens, where in Table 2 we already recognized that D & B counted less employment than the Census.

TABLE A-1

NEW YORK SMSA ESTABLISHMENT/EMPLOYMENT DISTRIBUTIONS BY INDUSTRY

Industry	Employment 1967 in thousands				Establishments 1967			
	Census	% Total	D & B	% Total	Census	Total	D & B	% Total
Food	58.7	5.8%	64.1	5.6%	1.245	3.7%	1,240	3.2%
Tobacco	–	–	1.8	0.2	17	0.0	40	0.1
Textiles	37.7	3.7	54.0	4.7	1.360	1.0	2.075	5.3
Apparel	260.4	25.5	220.2	19.2	10,549	31.4	9,838	25.5
Lumber	6.5	0.6	9.1	0.8	500	1.5	562	1.4
Furniture	20.6	2.0	28.2	2.5	1,206	3.6	1,682	4.3
Paper	26.6	2.6	40.6	3.5	646	1.9	872	2.2
Printing	153.7	15.1	150.2	13.1	4,973	14.8	5,399	13.9
Chemicals	33.0	3.2	77.6	6.8	890	2.6	1,747	4.5
Petroleum	0.8	0.1	1.8	0.2	45	0.1	100	0.3
Rubber	17.8	1.7	20.4	1.8	616	1.8	779	2.0
Leather	31.4	3.1	35.7	3.1	876	2.6	1,087	2.8
Stone	11.2	1.1	11.3	1.0	540	1.6	688	1.8
Primary Metals	13.7	1.3	18.2	1.6	293	0.9	456	1.2
Fabricated Metals	53.8	5.3	63.2	5.5	2,191	6.5	2,430	6.2
Machinery	39.9	3.9	54.6	4.8	1,871	5.6	2,306	5.9
Electrical Mach.	94.2	9.2	104.5	9.1	1,254	3.7	1,883	4.8
Trans. Equip.	56.4	5.5	56.0[1]	4.9	271	0.8	327	0.8
Instruments	26.8	2.6	48.5	4.2	665	2.0	998	2.6
Misc.	76.7	7.5	87.8	7.6	3,603	10.7	4,336	11.1
	1,019.9	100.0%	1,147.8	100.0%	33,614	100.0%	38,845	100.0%

[1] Corrected figures.

TABLE A-2

ESTABLISHMENT DISTRIBUTIONS

County	Census	% Total	Dun & Bradstreet[1]	% Total[2]
Brooklyn	6,384	18.5%	6,510 (867)	17.2%
Manhattan	17,841	51.8	21,608 (3,591)	54.9
Queens	2,972	8.6	3,181 (452)	8.3
Richmond	176	0.5	104 (20)	0.3
Bronx	1,748	5.1	1,654 (198)	4.4
Suffolk	1,475	4.3	1,043 (200)	2.6
Nassau	2,271	6.6	2,681 (376)	7.0
Rockland	196	0.6	373 (75)	0.9
Westchester	1,362	4.0	1,691 (285)	4.3
SMSA Total	34,425	100.0%	38,845 (6,064)	100.0%

[1] Establishments with zero employment in brackets below total. Census does not count such establishments.

[2] Net of establishments with zero employment to be consistent with the census data.

TABLE A-3

LARGE ESTABLISHMENTS BY COUNTY

County	Census[1]	D & B
Manhattan	200	234
Brooklyn	102	106
Queens	101	95
Bronx	25	29
Richmond	6	7
Nassau	55	72
Westchester	52	59
Suffolk	26	31
Rockland	10	20
	575	653

[1] Census figures include all establishments employing more than 250 persons, while the D & B figures count establishments employing 240 or more.

The census reports size distributions of establishments by 2 digit industry within county and it would have been desirable to make the comparison with similar D & B data. However, the census reports these figures of net of central and administrative office establishments while the D & B designation is headquarters and detached office establishments. Because the two categorizations differ significantly it was impossible to make this desired comparison. Even without this comparison, however, we remain satisfied that the Dun & Bradstreet DMI file is accurate for the purposes of location analysis.