1976

An Error in Estimating Urban Density Functions Using Census Tract Data

Mark W. Frankena

Follow this and additional works at: https://ir.lib.uwo.ca/economicsipsle_dp

Part of the Economics Commons

Citation of this paper:
Discussion Paper 013

AN ERROR IN ESTIMATING URBAN DENSITY FUNCTIONS USING CENSUS TRACT DATA

Mark W. Frankena

RESEARCH PROGRAM:
IMPACT OF THE PUBLIC SECTOR ON LOCAL ECONOMIES

Department of Economics
The University of Western Ontario
London Ontario Canada
Discussion Paper 013

AN ERROR IN ESTIMATING URBAN DENSITY FUNCTIONS USING CENSUS TRACT DATA

Mark W. Frankena

October 1976
AN ERROR IN ESTIMATING URBAN DENSITY
FUNCTIONS USING CENSUS TRACT DATA

Mark W. Frankena

I. Introduction

The purpose of this paper is to point out that there is an error in a number of published estimates of urban population density functions based on census tract data. The essence of the problem is that raw census tract data on population densities constitute a biased sample of observations on densities in an urban area. If this bias is overlooked, as it is in recent articles by Kau and Lee [3] and McDonald and Bowman [4] and in Muth's classic study [5, 141-145], estimates of density functions based on census tract data will overstate the expected density at each distance from the centre of the urban area. Fortunately, once the bias in the sample is recognized, the census tract observations can be weighted in a manner which offsets the non-randomness of the sample and leads to unbiased estimates of the density functions.

II. Model of Urban Population Density and Census Tract Mapping

Suppose that an urban area is divided into a large number of equal size "districts" which are mutually exclusive and exhaustive. We hypothesize that the true population density function for this urban area (excluding the central business district) has the form:

\[ Y_i = \alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2 + \alpha_3 x_i^3 + \epsilon_i \]  \hspace{1cm} (1)
where $Y_i =$ population density of district $i$, $X_i =$ distance of district $i$ from the centre of the urban area, and $\varepsilon_i$ is a stochastic error term which is normally distributed with zero mean and constant variance.

Suppose now that census tracts are formed by aggregating districts which are the same distance from the centre according to the following rules: (a) the districts consolidated have the same density, i.e., they have the same error terms in equation (1); and (b) the populations of each of the resulting census tracts should lie within a narrow range, which implies that census tracts formed by consolidating densely populated districts will be smaller than census tracts formed by consolidating sparsely populated districts. These are, in fact, two of the rules used in delineating census tracts by Statistics Canada.\[1\]

The following criteria were used to delineate these areas: (1) a population between 2,500 and 8,000 ...; (2) an area as homogeneous as possible in terms of economic status and living conditions; ... [6, 95-721, 3]

While the stochastic error terms for the densities of equal sized districts would be distributed with zero mean and constant variance in the probability limit, this would obviously not be the case for the error terms for the densities of the new set of census tracts. Since each census tract formed by consolidating sparsely populated districts (i.e., districts with large negative $\varepsilon_i$'s) will consist of a larger number of districts than each census tract formed by consolidating densely populated districts (i.e., districts with large positive $\varepsilon_i$'s), the distribution of the error terms for the densities of the census tracts would have a positive mean and would
be skewed to the left in the probability limit. In other words, because the sparsely populated districts are consolidated to a greater degree than the densely populated districts, the sparsely populated districts will be underrepresented in the census tract data.

If equation (1) were estimated using census tract data, there would be an upward bias in the estimate of $\alpha_0$ because of the positive mean of the error term in the probability limit. In other words, the estimated population density function would lie above the true function in the probability limit because districts with negative error terms are underrepresented relative to districts with positive error terms in the census tract sample.

III. Correcting for Sampling Bias

It should be clear from the preceding discussion that the basic problem in using raw census tract data is the underrepresentation of low density districts relative to high density districts at a given distance from the centre. Fortunately, by a simple weighting of the census tract data one can still obtain an unbiased estimate of the density function. What is required is to weight census tract observations in proportion to their areas. Suppose census tract A has twice the area of census tract B, i.e., census tract A contains twice as many districts as census tract B. If the observation for census tract A is given twice the weight of census tract B in the estimation, then effectively the individual districts in each census tract will be given equal weight.

Thus, in order to estimate equation (1) using census tract data (for either all census tracts or a random sample of census tracts [$j=1,...,n$] in an urban area), one should select $\hat{\alpha}_0$, $\hat{\alpha}_1$, $\hat{\alpha}_2$, and $\hat{\alpha}_3$ to minimize the
following weighted sum of squared residuals:

$$
\sum_{j} a_j (y_j - \hat{\alpha}_0 - \hat{\alpha}_1 x_j - \hat{\alpha}_2 x_j^2 - \hat{\alpha}_3 x_j^3)^2
$$

where $a_j$ is the area of census tract $j$.

IV. Example

To demonstrate the empirical significance of the estimation problem being discussed, we used two methods to estimate equation (1) using 1971 census tract data for Toronto. First, we used ordinary least squares with each census tract observation weighted equally regardless of area ("unweighted estimate"). Second, we used ordinary least squares with each census tract observation weighted by area, as suggested in equation (2) ("weighted estimate"). In each case, we used all census tracts within the political jurisdiction of Metropolitan Toronto (C.T. 1 to 378) excluding the central business district (C.T. 1-3, 8-20, 61-64, 88-89), rather than all census tracts in the Census Metropolitan Area of Toronto. Published data on density and area of census tracts were both obtained from the 1971 Census of Canada [6, 98-701], while data on distances were measured by hand from the centre of each census tract to the intersection of King Street and Yonge Street using census tract maps.²

The two estimates of the density function are presented in Table 1, and the functions are plotted in Figure 1 for distances between 1.39 and 12.40 miles from the centre. It will be observed that, as expected, the "unweighted estimate" lies everywhere above the "weighted estimate". It is our contention that the difference is a result of the sampling bias inherent in the unweighted approach, given the rules that are employed in
TABLE 1
Estimates of Equation (1)

<table>
<thead>
<tr>
<th></th>
<th>Unweighted Estimate</th>
<th>Weighted Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\alpha}_0$</td>
<td>58141.7 (14.52)</td>
<td>54144.8 (14.82)</td>
</tr>
<tr>
<td>$\hat{\alpha}_1$</td>
<td>-14574.2 (-6.35)</td>
<td>-15071.0 (-8.78)</td>
</tr>
<tr>
<td>$\hat{\alpha}_2$</td>
<td>1449.13 (3.79)</td>
<td>1651.32 (6.81)</td>
</tr>
<tr>
<td>$\hat{\alpha}_3$</td>
<td>-50.0212 (-2.62)</td>
<td>-62.4071 (-6.00)</td>
</tr>
</tbody>
</table>

$t$-statistics in parentheses

*The parameters of the "weighted estimate" were computed by estimating the following equation by ordinary least squares:

$$\sqrt{\bar{a}_j} y_j = \alpha_0 \sqrt{\bar{a}_j} + \alpha_1 \sqrt{\bar{a}_j} x_j + \alpha_2 \sqrt{\bar{a}_j} x_j^2 + \alpha_3 \sqrt{\bar{a}_j} x_j^3 + \sqrt{\bar{a}_j} \epsilon_j$$

In a model such as this which has a zero intercept, $R^2$ may turn out to be negative or greater than one, and it cannot be used as a measure of goodness of fit [1, 85-90].
FIGURE 1
Estimates of Equation (1)

$Y_1$ ('000 people per square mile)

"unweighted estimate"

"weighted estimate"

$Y_1$ (miles)
delineating census tracts.

V. Test of the Reasonableness of the Estimates

McDonald and Bowman [4, 248] propose that "as pointed out by Clark [2] in the classic paper on urban density functions, a reasonably straightforward test of the reliability of a population density function is to predict total population by computing the integral of the function within the appropriate limits." McDonald and Bowman were disturbed to find that their estimates of density functions and the estimates made by Muth both lead to systematic overstatements of the total populations of the urban areas involved. Typically, these overstatements are on the order of 30 per cent for McDonald and Bowman's density functions. Moreover, this overstatement occurs for all functional forms and even when one uses a 100 per cent sample of census tracts.

It is the contention of this paper that the explanation for this overprediction is that the density functions estimated by McDonald and Bowman and by Muth systematically overstate the average population density at each distance from the centre of the city, in the same way that the "unweighted estimate" in Figure 1 does.

To demonstrate this, we computed the integrals of the two functions in Table 1 over a close approximation of the area of Metropolitan Toronto excluding the central business district (CBD). We used a semicircular ring (since Toronto lies next to a lake) with inner radius of 1.39 miles and outer radius of 12.40 miles. The true 1971 population of Metropolitan Toronto excluding the CBD was 2,042,323. The integral of the "unweighted estimate" of the density function was 2,498,400, or 22.3 per cent above
the true population. By contrast, the integral of the "weighted estimate" was 2,027,000, or within one per cent of the true population.
FOOTNOTES

*Associate Professor of Economics. I am grateful to David F. Burgess, Gordon W. Davies, Erik F. Haites, Arthur J. Robson, David T. Scheffman, and Robert S. Woodward for suggestions and to Linda Newton for research assistance.

1The same is true for the U.S. Bureau of the Census which states that "census tracts are small areas, having a population generally between 3,000 and 6,000 .... The tract areas are established with a view to approximate uniformity in population, with some consideration of uniformity in size of area .... Each tract is designed to include an area fairly homogeneous in population characteristics." [7, 36].

2Small measurement errors would arise in the distance variable because the centre of each census tract was determined by eye. Also, a small bias in the coefficients could result from the fact that the entire census tract is assumed to be located at the same distance from the centre of the urban area as the centre of the census tract.

3The outer radius is the radius of a semicircle with the same area as Metropolitan Toronto including the portion of the CBD lying north of King Street. The inner radius is the radius of a semicircle with the same area as that portion of the CBD lying north of King Street. The area of the CBD lying south of King Street, including the railways, expressways, harbour, and islands, was omitted from all aspects of the present study.
REFERENCES


RESEARCH PROGRAM:
IMPACT OF THE PUBLIC
SECTOR ON LOCAL ECONOMIES

The Department of Economics, University of Western Ontario, has recently mounted a long-term research program on the Impact of the Public Sector on Local Economies. The program publishes Discussion Papers and Research Studies. Copies may be obtained by writing:

Mrs. Jayne Dewar
Secretary
Urban Research Program
Department of Economics
The University of Western Ontario
London, Ontario N6A 5C2
Canada

The following Discussion Papers and Research Studies are currently available:

DISCUSSION PAPERS:

* 001 P. T. Chinloy, "Hedonic Prices and Age Deterioration in Urban and Suburban Housing" (September 1975). 30 pp. - $1.50

002 M. W. Frankena, "Alternative Models of Rent Control" (September 1975).


004 A. J. Robson, "The Effect of Urban Structure on Ambient Pollution" (October 1975).


007 M. F. Goodchild and P. J. Booth, "Modelling Human Spatial Behavior in Urban Recreation Facility Site Location" (January 1976).
DISCUSSION PAPERS: (continued)


010 P. T. Chinloy, "Depreciation, Adverse Selection and Housing Markets" (May 1976).


012 A. J. Robson, "Income Uncertainty and Urban Location" (September 1976).


RESEARCH STUDIES:

* 01 G. W. Davies and P. L. Jackson, "A Model of the Urban Housing and Residential Land Markets" (September 1975). 132 pp. - $6.60

02 G. P. Schaefer, "The Urban Area Production Function and the Urban Hierarchy: The Case of Saskatchewan" (October 1975).

---

*This report is out-of-print. Xerox copies can be obtained for a minimal charge of 5¢ per page.*