UCI-ITS-WP-91-5

The Distribution of Population in a Contemporary Metropolitan Area: The Case of Orange County [a replication]

UCI-ITS-WP-91-5

Shunfeng Song [1]
Kenneth A. Small [2]

[1] Department of Economics and
Institute of Transportation Studies, University of California, Irvine
Irvine, California 92697-3600, U.S.A.

June 1991

Institute of Transportation Studies
University of California, Irvine
Irvine, CA 92697-3600, U.S.A.
http://www.its.uci.edu

The Distribution of Population in a Contemporary Metropolitan Area: The Case of Orange County [A Replication]

Shunfeng Song

Department of Economics
University of California at Irvine
Irvine, CA 92717

Working Paper
June 1991

UCTC No. 44

The University of California Transportation Center
University of California at Berkeley

Summary

This paper examines the population distribution in a rapid development area - Orange County in 1980. The population density gradients are estimated with a polycentric model as well as with a monocentric model. The paper shows that the polycentric model of population distribution fits Orange County better than the monocentric model, and that the population in Orange County is more dispersed than in Los Angeles County. It is found that Downtown Los Angeles has a significant influence on the population distribution in Orange County. The aggregate influences of centers on population distribution have a wide range which are affected by the inclusion of center Downtown Los Angeles.

The Distribution of Population in a Contemporary Metropolitan Area: the Case of Orange County

[A Replication]

Shunfeng Song*

Introduction

Population distribution is a basic concern of urban spatial structure. Urban economists agree that population density and distance from center are negatively related. Until recently almost all density analysts have retained the assumption of a monocentric city, and the most common functional form for modelling the relationship of density-distance is the negative exponential (Clark, 1951; Muth, 1969; Kemper and Schmenner, 1974; Mills and Hamilton, 1984).

The monocentric model of population distribution, however, does not explain well the distribution of population in the Los Angeles area, and the same results may be expected in other metropolitan areas, particularly in the West (Gordon, Richardson and Wong, 1986). Using the census data on population in Los Angeles area for 1970 and 1980, Gordon *et al*. estimated the population density gradients with a polycentric model, and found that the polycentric model of population distribution fits Los Angeles much better than the monocentric model. This paper attempts to further their research to a rapid development urban area.

---

*I wish to thank Amihai Glazer and Kenneth A. Small for valuable discussions and comments.

The purpose of this paper is threefold. First, it is to estimate the population density gradients for a contemporary metropolitan area - Orange County in 1980, with a polycentric model and a monocentric model respectively. The fits of these two models in explaining the population distribution are to be compared. The second purpose is to examine the influence of Los Angeles County on the population distribution in Orange County. The last purpose of this paper is to investigate the aggregate influences of population centers on the whole Orange County area.

A Polycentric Model

A polycentric model is formulated which assumes that the population density at a given point results from the layering of influences generated by all centers (Griffith, 1981a; Griffith, 1981b; Gordon, Richardson and Wong, 1986). This implies that density at any location is the vertical sum of negative exponential density functions, each reflecting the influence of a center at that location.

Let N be the number of population centers and M be the number of zones in a metropolitan area. A polycentric version of the negative exponential form of the density function is formulated as a straightforward extension of the monocentric model:

$$D_m = \sum_{n=1}^{M} A_n e^{-b_n r_{mn}} + v_m, \qquad (m = 1, 2, \ldots, M)$$

where D_m = the observed population density in zone m,

r_{mn}= the distance between zone m to center n,

v_m = the error term of population density associated with zone

m. The first term on the right side of the equation is the predicted population density by the polycentric model. A_n, b_n are the parameters to be estimated for each center n. A_n is the intercept and b_n is the density gradient for center n respectively.

The polycentric model allow us to examine the absolute or the relative influence of each center on a given location. The absolute influence of center n on location m is measured by the term $A_n e^{-b_n r_{mn}}$, and the relative influence of center i to center j on location m is $(A_i e^{-b_i r_{mi}})/(A_j e^{-b_j r_{mj}})$, where A_n, A_i, A_j, b_n, b_i, and b_j are the estimates. Furthermore, we can calculate the aggregate influence of center n on the whole area, which is the sum of its influence on each zone m,

$$AINF = \sum_{m=1}^{M} (A_n e^{-b_n r_{mn}}) S_m ,$$

where S_m is the area of zone m.

The polycentric model also allows us to consider the influences of centers outside the metropolitan area studied. In fact, these centers can be simply treated as additional centers in the model, and we let the number of centers be $N+K$ ($N+K$ should be smaller than $M/2$ because each center has two parameters to be estimated), where K is the number of centers considered outside the area studied. This consideration is plausible when we examine the geographic distribution of population density for a sub-area within a large metropolitan area. For example, it is important to consider the influence of Downtown Los Angeles on the population distribution in Orange County while examining the distribution of population in Orange County area.

Study Area and Data

Orange county, the nation's 17th largest metropolitan area named "Anaheim - Santa Ana - Garden Grove" with 780 square miles, is chosen as the study area. Orange County experienced high growth since 1950. It had a population of 200,000 in 1950. By 1960 its population grew to 750,000. Between 1960 to 1970, the most rapid expansion period, the population doubled. Growth continued in the 1970's with a less rapid rate and 500,000 people were added (Source: U.S. Census 1960,70,80). It has become the nation's sixth most populous county. It has no dominant central city, and is economically independent from central Los Angeles area, providing 84 jobs for every 100 resident labor-force members (Berechman and Small, 1989). The traditional monocentric model is thus not suitable to describe the geographic distribution of population density in this area.

Data on the 1980 population for 193 Analysis Zones (defined by the Southern California Association of Governments, SCAG) in Orange county are used. Analysis zones are aggregates of census blocks. They are larger in average geographic size than census tracts, divided on the land use pattern and the road network. Therefore, they are better suited for transportation planning. The data on the area for these zones measured in acres are also available from the SCAG, which are used to compute the population densities. Distances are measured by the air-line distance between zones connecting the central point in each zone via a zone map.

Estimation Methods

Quasi-population-centers are identified by the following procedures. First, zones which have higher population densities than each of their contiguous zones are sorted, and treated as population center candidates (similar to McDonald's approach, 1987). It turns out that there are 21 center candidates in Orange County area. Second, for each of these candidates, density differences between the candidate and its contiguous zones are calculated. Then a difference ratio (percent change) for each contiguous zone is computed, which equals the difference divided by the candidate's density. Third, a candidate is defined as a quasi-center if it satisfies the following two criteria: (1) its density is at least 10 percent higher than the density of each contiguous zone, (2) its density is not 50 percent higher than all contiguous zones. If a candidate has a density nearly the same as some of its contiguous zones, it can be better identified as a center by including its contiguous zones, and the combined area is treated as a second-round center candidate. Here, a candidate is combined with its contiguous zones which have densities higher than 90 percent of its density. The density of this combined area is the ratio of combined population with combined area. On the other hand, if a candidate has a much higher density than all its contiguous zones, it implies that the candidate is a purely local population density peak which does not have significant effect on whole area. They are eliminated as population centers. Fourth, step 3 is repeated for the second-round candidates until they can be either accepted or rejected as quasi-centers.

Based on the above procedures, 12 quasi-centers are identified in Orange County. Table 1 describes them. Column 1 shows the locations of quasi-centers, and column 2 presents their population densities.

The polycentric density gradients are estimated by minimizing the nonlinear least squares with the quasi-centers identified. A quasi-center is not a population center if its density gradient estimate (negative) is very large, because large negative density gradient implies that the quasi-center has little influence on other zones. Hence, the population centers are semi-endogenously identified. In order to investigate the influence from outside Orange county, a center in Los Angeles County (Downtown Los Angeles) is also included.

Denote SSR to be the sum of squared residuals. Then

$$SSR = \sum_{m=1}^{M} v_m^2 = \sum_{m=1}^{M} (D_m - \sum_{n=1}^{N} A_n e^{-b_n r_{mn}})^2 ,$$

The polycentric gradients b_n and intercepts A_n are estimated to minimize SSR.

Estimation Results

The polycentric density gradients with quasi-centers are first estimated. four quasi-centers (La Habra, Seal Beach, Corona del Mar, and East Irvine) are found to have large negative gradients, implying they are purely local population density peak and thus have little influences on other zones. They are eliminated. The remaining eight quasi-centers are population centers.

Then the polycentric gradients are re-estimated with the population centers. Table 2 presents the results of estimation. Column 1 is the locations of the centers. Column 2 and 3 present the estimates with their t-statistic values in parentheses. Among eight intercept estimates, only one is insignificant at 5 percent level. Half of the gradient estimates are significant at 5 percent level.

The monocentric density gradient is estimated as well. The estimates are summarized in Table 3. Followed Alperovich's method (1982), the moncentric center is chosen endogenously among the quasi-population-centers such that the center has a highest R^2 value, which shows how well the estimated model fits the actual population distribution. The center Midway City has highest R^2 value, thus it is chosen as the mono-center in Orange County. In fact, Midway City is dense with low income population. The center Downtown Los Angeles is not identified as the mono-center, indicating that the main influence on population distribution is not from Downtown Los Angeles.

Since the mono-center is within the Orange County itself, it is more direct to compare the monocentric model with this mono-center and the polycentric model with centers within Orange County. Consider first the fits of models in explaining the population distribution. The monocentric model, although it has high t-value estimates, has a much lower R^2 value. It explains only 40 per cent of population distribution while the polycentric model can explain 63 per cent of the population distribution. The R^2s, however, are not directly comparable because of the differences in the number of variables in the two models. Hence, the *F*-test (please see the appendix for detailed discussion) is used to test the significance of adding additional centers into the model (Gallant, 1975). The result of *F*-test shows that the polycentric model is statistically superior at 5 percent significant level. Thus the polycentric model explains the population distribution in Orange County much better.

To examine whether the population is more dispersed in Orange County than in Los Angeles County, the density gradient estimates for these two counties are compared (Table 2 vs Table 5). Since each polycentric density gradient describes the percentage fall in population density for a unit increase in distance from each population center, the density gradient estimates measures the dispersion of population in the whole area. Smaller negative density gradients imply more dispersion of population distribution. Comparing the estimates of density gradients for Orange County with the estimates by Gordon *et al.* for Los Angeles County, the population in Orange County is indeed more dispersed, because its polycentric density gradient estimates are almost always smaller than the estimates for Los Angeles County.

The influence on population distribution from outside Orange County can be measured to some degree by including an outside main population center in the polycentric model. Table 4 presents the estimates when the polycentric model includes the center Downtown Los Angeles. The results of estimation (Table 2 and Table 4) show that the polycentric model can better explain the population distribution in Orange County by adding the center Downtown Los Angeles. The polycentric model with center Downtown Los Angeles explains 66 per cent of population distribution. The polycentric model without center Downtown Los Angeles, however, explains 63 per cent of population distribution. The *F*-test confirms that Downtown Los Angeles has a statistically significant (at 5 percent level) influence on the population distribution in Orange County. Therefore, including main outside population center(s) can improve the fit of polycentric model in explaining the population distribution.

The predicted aggregate influence on the whole studied area is also calculated for each center. Column 4 in both Table 2 and Table 4 report the results with their approximate *t*-values in parentheses. Two interesting points appear here. (1) The aggregate influences have a wide range among centers, implying that some centers are more important than others. The *t*-values of *AINF* in Table 2 and Table 4 show that three of the four largest *AINF* are statistically significant at 5 percent level. (2) The aggregate influences depend on whether the polycentric model includes the center Downtown Los Angeles. With the center Downtown Los Angeles, the largest four centers are South Fairview, Midway City, El Toro and Downtown Los Angeles. The four largest centers become Stanton, Midway City, El Toro and Santa Ana if the polycentric model excludes the center Downtown Los Angeles. The aggregate influence of center Stanton has a big jump in these two cases, from the smallest center to the largest center. It is obvious that center Stanton is shadowed by the inclusion of center Downtown Los

Angeles, indicating that the center Downtown Los Angeles does affect the population distribution in Orange County.

Conclusion

A polycentric density gradient algorithm has been applied in this paper to estimate the population density gradients in Orange County. The monocentric density gradient is also estimated. Four main conclusions can be made. First, the polycentric model fits the population distribution in Orange County much better than the monocentric model. This conclusion is consistent with the fact that there is no dominant center in Orange County. Second, the population distribution in Orange County is more dispersed than in Los Angeles County. Comparing the polycentric density gradients between these two counties, Orange County has smaller negative gradient estimates, implying that its population is more dispersed. Third, the influence on population distribution from outside Orange County is significant. A polycentric model can better explain the population distribution when it includes some main center(s) in Los Angeles County. Finally, this paper shows that the aggregate influences of different centers have a wide range, and they are affected by the inclusion of center Downtown Los Angeles, indicating that some centers are more important than others on population distribution, and Downtown Los Angeles has a significant influence on the population distribution in Orange County.

Appendix: F-test

The nonlinear regression model can be presented by the equations

$$D_m = f(R_m, B) + e_m \qquad m = 1, 2, \ldots, M,$$

where $R_m = (r_{m1}, r_{m2}, \ldots, r_{mN})$, B is a p-dimensional vector of unknown parameters, and the e_m represent unobservable errors.

Nonlinear least square estimation is to estimate the unknown parameters by minimizing the sum of squared residuals, i.e. to estimate B by minimizing

$$SSR(B) = \sum_{m=1}^{M} [D_m - f(R_m, B)]^2 .$$

The F-test is used to test a hypothesis of restrictions on some unknown parameters,

$$H_0: t = t_0 \quad \text{against} \quad H_1: t = t_0$$

where t is a q-dimensional subvector of the parameter vector B.

One way to test the hypothesis is to compare the minimal $SSRs$ (Gallant, 1975). Assume SSR^u to be the minimal SSR when there are no restrictions on parameters, and SSR^r to be the minimal SSR with restrictions on parameters. Then $(SSR^r - SSR^u)(M-p)/(qSSR^u)$ is F-distributed. The null hypothesis can be rejected at significance level a if the value of this ratio is bigger than F_α. Equivalently, we can reject the null hypothesis if the value of SSR^r/SSR^u exceeds a critical value C^*, where C^* is defined as

$$C^* = 1 + qF_\alpha/(M-p),$$

where F_α is the upper 100α percentage point of a central

F-distribution with degree of freedom $(q, M\text{-}p)$.

Monocentric model vs polycentric model: We have a monocentric model when the parameters in polycentric model are set to be zeros for all centers but the mono-center. Note that the mono-center (Midway City) is chosen to have a highest R^2 value among the centers. Thus it has a lowest value of SSR^r/SSR^u. To test whether the monocentric model is significantly better in explaining the population distribution (H_0: the monocentric model is superior), I use the above F-test with p=16, q=14, and M=186. There are 8 centers within Orange County and hence the polycentric model has 16 parameters (p=16). q=14 because only two parameters appear the monocentric model. M is the observation number(sample size), which equals to 186. The critical value of C^* for significant level α=0.05 is

$$C^* = 1 + (14*F_{\alpha=0.05})/(186-16) = 1.1375.$$

The value of SSR^r/SSR^u with the mono-center is 1.6191. Thus, $SSR^r/SSR^u > C^*$ for the mono-center. Therefore, I reject the null hypothesis, implying the polycentric model is superior.

Influence from center Downtown Los Angeles: Since I include an additional center Downtown Los Angeles into the polycentric model, p=18. To test whether the influence from center Downtown Los Angeles, I test the null hypothesis H_0: the parameters for center Dowtown Los Angeles are zeros. I calculated SSR^r/SSR^u = 1.0796. The critical value C^* = 1 + (2*$F_{\alpha=0.05}$)/(186-18) = 1.0357, which is smaller than 1.0796. Thus I reject the null hypothesis. Therefore, center Downtown Los Angeles has significant influence on the population distribution in Orange County.

References

Alperovich, G. (1982). "Density Gradients and the Identification of the Central Business District." *Urban Studies*, Vol. 17: 313-320.

Berechman, J. and Small, K. A. (1988). "Modeling Land Use and Transportation: An Interpretive Review for Growth Areas." *Environment and Planning A*, Vol. 20: 1285-1309.

Clark, C. (1951). "Urban Population Densities." *Journal of Royal Statistical Soceity*, Series A, Vol. 114: 490-496.

Gallant, A. R. (1975). "Nonlinear Regression." *The American Statistician*, Vol. 29: 73-81.

Gordon, P., Richardson, H. W. and Wong, H. L. (1986). "The Distribution of Population and Employment in a Polycentric City: The Case of Los Angeles." *Environment and Planning A*, Vol. 18: 161-173.

Griffith, D. A. (1981a). "Evaluating the Transformation from a Monocentric to a Polycentric City." *Professional Geographer*, Vol. 33: 189-196.

Griffith, D. A. (1981b). "Modeling Urban Population Density in a Multi-Centered City." *Journal of Urban Economics*, Vol. 9: 298-310.

Kemper, P. and Schmenner, R. (1974). "The Density Gradient for Manufacturing." *Journal of Urban Economics*, Vol. 1: 410-427.

McDonald, J. F. (1987). "The Identification of Urban Employment Subcenters." *Journal of Urban Economics*, Vol. 21: 242-258.

Mills, E. and Hamiltion, B. (1982). *Urban Economics*, 3rd ed. Scott Foresman, Glenview, Illinois.

Muth, R. (1969). *Cities and Housing*, University of Chicago Press, Chicago.

Table 1: Quasi-Population-Centers

Center	Density(persons/mile^2)
La Habra	7177
Placentia	8394
Anaheim	9166
Stanton	9017
Seal Beach	9501
Midway city	10236
South Fairview	8458
Corona del Mar	6352
Orange	6768
Santa Ana	13510
El Toro	7908
East Irvine	5674

Table 2: Polycentric Estimates without Center Downtown L.A.

Center	Intercept	Gradient	AINF
Placentia	4812.0 (2.5736)	-1.20770 (-1.6073)	27444 (0.8971)
Anaheim	3663.9 (2.0545)	-0.61083 (-1.2054)	63360 (0.7127)
Stanton	5563.0 (2.4912)	-0.07830 (-3.5222)	1639553 (2.8666)
Midway City	4412.6 (3.0467)	-0.18614 (-1.4378)	479148 (0.9152)
South Fairview	4338.6 (2.8849)	-0.49075 (-1.9121)	96817 (1.2119)
Orange	2008.2 (1.0571)	-3.10990 (-0.2338)	3736 (0.3554)
Santa Ana	10056.0 (5.7642)	-0.74725 (-4.6564)	131438 (2.8036)
El Toro	7467.9 (4.1848)	-0.64220 (-3.072)	372658 (3.1323)

$R^2 = 0.63$

t-values are in parentheses.

Table 3: Monocentric Estimates

Mono-Center	Intercept	Gradient
Midway City	9111.5 (16.427)	-0.068073 (-8.9107)
$R^2 = 0.40$		

t-values are in parentheses.

Table 4: Polycentric Estimates with Center Downtown L.A.

Center	Intercept	Gradient	AINF
Placentia	6179.7 (3.8807)	-0.53726 (-2.2637)	146870 (1.4742)
Anaheim	5065.2 (2.7835)	-0.59625 (-1.5920)	91241 (0.9030)
Stanton	3719.6 (2.2252)	-0.50229 (-1.2280)	88374 (0.6370)
Midway City	6569.2 (4.7776)	-0.25068 (-2.3476)	463369 (1.5440)
South Fairview	4825.2 (4.4514)	-0.12036 (-3.3211)	993885 (2.9718)
Orange	2611.4 (1.8185)	-0.45100 (-0.96703)	101801 (0.6259)
Santa Ana	10234.0 (6.1446)	-0.66009 (-3.9097)	169907 (2.2543)
El Toro	7687.3 (4.5020)	-0.60708 (-3.0644)	414639 (3.1212)
Downtown L.A.	1.446E+08 (0.26763)	-0.38126 (-2.1489)	244303 (1.8663)

R^2 = 0.66

t-values are in parentheses.

Table 5. Los Angeles County population density gradients, 1980

Area ID#*	Intercept	Gradient
1	21800	-1.05
2	4500	-0.42
7	8630	-1.48
8	8610	-0.55
9	14800	-1.11
11	5390	-1.89
14	11400	-1.08
15	13300	-1.63
16	14100	-1.72
17	15900	-1.44
18	24100	-1.70
19	33400	-2.10
20	20000	-10.50
21	7240	-0.48
22	7560	-0.90
23	23400	-1.56
24	16700	-4.85
25	30000	-2.10
26	14600	-1.49
27	32900	-1.10
29	36300	-1.66
30	9150	-2.73
31	21700	-1.72
33	10500	-0.60
34	12900	-1.86

*Identification number for the area in the census.

Source: Table 3 in Gordon, Richardson and Wong, page 165, 1986.